

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

VYHODNOCOVÁNÍ KVALITY ŘEČOVÉHO SIGNÁLU

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

PETER TULEJA

BRNO 2011



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



**FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ**
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

VYHODNOCOVÁNÍ KVALITY ŘEČOVÉHO SIGNÁLU

ASSESSMENT OF SPEECH SIGNAL QUALITY

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

PETER TULEJA

VEDOUCÍ PRÁCE
SUPERVISOR

Ing. IVAN MÍČA

BRNO 2011



**VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ**

**Fakulta elektrotechniky
a komunikačních technologií**

Ústav telekomunikací

Bakalářská práce

bakalářský studijní obor
Teleinformatika

Student: Peter Tuleja

ID: 115301

Ročník: 3

Akademický rok: 2010/2011

NÁZEV TÉMATU:

Vyhodnocování kvality řečového signálu

POKYNY PRO VYPRACOVÁNÍ:

Seznamte se s problematikou vyhodnocování kvality řečového signálu subjektivně i objektivně. Vyberte jednoduchou subjektivní metodu, s jejíž pomocí vytvoříte malou databázi subjektivně hodnocených nahrávek. Navrhněte a implementujte rámcový systém pro objektivní vyhodnocování kvality řeči a pro porovnání výstupů objektivních metod s referenčními, subjektivně ohodnocenými nahrávkami v databázi. Součástí systému má být knihovna objektivních metod.

DOPORUČENÁ LITERATURA:

- [1] PSUTKA, J., et al. Mluvíme s počítačem česky. Praha: Academia, 2006. 752 s. ISBN 80-200-1209-1.
- [2] LOIZOU, P. C. Speech Enhancement. CRC Press, 2007. ISBN 978-0849350320.

Termín zadání: 7.2.2011

Termín odevzdání: 2.6.2011

Vedoucí práce: Ing. Ivan Míča

prof. Ing. Kamil Vrba, CSc.

Předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Táto práca pojednáva o metódach vyhodnocovania kvality rečového signálu. V krátkosti popisuje subjektívne metódy pre určenie kvality rečového signálu. Zo subjektívnych metód je prezentovaná metóda párového porovnávania a MOS skóre. Podrobnejšie sú opísané intruzívne objektívne metódy - konkrétne metódy segmentálnej miery SNR v časovej doméne, metóda segmentálnej miery SNR vo frekvenčnej doméne a metóda normalizácie rámcov využívajúca odhad LSE. V závere je popísaný experiment, v ktorom sú spomínané metódy porovnané a štatisticky vyhodnotené.

KĽÚČOVÉ SLOVÁ

kvalita, reč, rečový signál, vyhodnotenie, objektívne, subjektívne, párové porovnávanie, komparatívny prístup, MOS skóre, SNR

ABSTRACT

This paper discusses methods for evaluating the quality of the speech signal. Briefly describe the subjective methods for determining the quality of the speech signal. From subjective methods the pairwise comparison and MOS score are presented. Objective intrusive methods are described in more detailed way - namely methods of the segmental SNR evaluated in time domain, method of the segmental SNR evaluated in the frequency domain and frame normalization method which uses LSE based estimator. At the end of this paper is described an experiment, in which the aforementioned methods are compared and then statistically evaluated.

KEYWORDS

quality, speech, speech signal, evaluation, objective, subjective, pairwise comparison, comparative approach, MOS score, SNR

TULEJA, Peter *Vyhodnocování kvality řečového signálu*: bakalářská práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2010. 56 s. Vedoucí práce byl Ing. Ivan Míča

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Vyhodnocování kvality řečového signálu“ jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

Brno

.....

(podpis autora)

Podakovanie

V prvom rade chcem poďakovať svojmu konzultantovi Ing. Ivanovi Míčovi, za jeho cenné rady, odbornú a metodickú pomoc pri vytváraní tejto práce. Ďalej chcem poďakovať Mgr. Jurajovi Tulejovi za jazykovú a štylistickú korekciu práce. Ďalšie poďakovanie patrí ľuďom, ktorí sa zúčastnili nahrávania databázy rečových vzoriek. Menovite (v abecednom poradí): Běleja Martin, Durkáčová Katarína, Kačmárová Katarína, Pružinský Ján, Trávníček Michal, Vránová Marta. Ďalšie poďakovanie patrí poslucháčom, ktorí sa podieľali na subjektívnych testoch. Menovite (v abecednom poradí): Bednár Jakub, Holíš Michal, Hradiský Marek, Jacko Róbert, Jakubec Ján, Jančuš Rastislav, Ludva Ondřej, Novák Daniel, Piro Šimon, Plačko Michal, Popovec Juraj, Pružinský Ján, Truhlářová Eva.

OBSAH

Úvod	10
1 Reč – vznik, prenos, vnímanie	11
1.1 Kvalita a zrozumiteľnosť	12
1.2 Vplyvy znižujúce kvalitu a zrozumiteľnosť	13
1.3 Algoritmy pre potláčanie šumu	14
1.4 Subjektívne metódy	15
1.4.1 Metóda párového porovnávania	15
1.4.2 MOS skóre	16
1.5 Objektívne metódy	17
1.5.1 Intruzívne metódy	18
1.5.2 Neintruzívne metódy	22
2 Praktická časť práce	23
2.1 Tvorba databázy nahrávok	23
2.2 Psychoakustický experiment	24
2.2.1 Návrh	24
2.2.2 Výsledky	28
2.3 Objektívne metódy	34
2.3.1 Implementácia	34
2.3.2 Výsledky	35
2.4 Vyhodnotenie objektívnych metód	37
Literatúra	43
Zoznam symbolov, veličín a skratiek	44
A Výsledky párového porovnávania	46
B Zdrojové kódy	50
B.1 Main.m	50
B.2 Framework.m	51

B.3	segSNR.m	52
B.4	FreqSNR.m	53
B.5	corelation.m	54
B.6	framenorm.m	55

ZOZNAM TABULIEK

1.1	Tabuľka porovnania kvality piatich podnetov metódou párového porovnávania	16
1.2	Tabuľka hodnotenia podnetov pomocou skóre MOS	17
1.3	Tabuľka váh pre centrálné kmitočty filtrov	20
1.4	Tabuľka parametrov a váh využitých pri výpočte R_{str}	22
2.1	Zoznam použitých nahrávok	24
2.2	Poradie dvojíc nahrávok pri párovom porovnávaní	25
2.3	Systém výberu šumov pre jednotlivých poslucháčov	28
2.4	Matica preferencií jednotlivca pre nahrávky obsahujúce šum z ul. Zubří .	29
2.5	Matica preferencií skupiny pre nahrávky obsahujúce šum z ul. Zubří .	30
2.6	Matica hodnôt $C_2(p_{jk})$ pre nahrávky obsahujúce šum z ul. Zubří . . .	31
2.7	Výsledky SNRseg	35
2.8	Výsledky SNRsegR	35
2.9	Výsledky fwSNRseg	36
2.10	Výsledky framenorm	36
A.1	Babble, pokus 1.	46
A.2	Factory, pokus 1.	46
A.3	White noise, pokus 1.	47
A.4	ul. Zubri, pokus 1.	47
A.5	Babble, pokus 2.	48
A.6	Factory, pokus 2.	48
A.7	White noise, pokus 2.	49
A.8	Zubri2, pokus 2.	49

ÚVOD

Tento dokument sa zaoberá problematikou vyhodnotenia kvality rečových signálov k posúdeniu výkonnosti algoritmov pre potláčanie šumu. V dokumente sú rozobraté spôsoby, akými sa dá kvalita vyhodnocovať a problémy, ktoré pri vyhodnocovaní nastávajú.

V prvej časti sa nachádza všeobecné a stručné oboznámenie s rečou a s úlohou algoritmov slúžiacich pre potlačenie šumu. V nasledujúcej časti je opísaný rámcový systém vyhodnocovania kvality rečového signálu. Prakticky to znamená proces vytvorenia nahrávky, jej spracovanie, normalizácia, strih, zašumenie, následné odšumenie použitím algoritmov pre potláčanie šumu. Rámcový systém ďalej zahŕňa návrh subjektívnej metódy, výber testovaných podnetov, výber pokusných osôb a štatistické overenie správnosti výsledkov subjektívnych testov. Ďalším bodom je použitie objektívnych intruívnych metód pre zistenie kvality rečového signálu (meranie matematickej „vzdialenosti“ spracovávaného signálu od referenčného). V závere je popísaný proces štatisticky správneho porovnania výsledkov a vyvodenie z toho vyplývajúcich záverov. Konkrétne boli použité dve subjektívne metódy a štyri objektívne metódy, ktoré sú implementované v programe Matlab. Výsledky jednotlivých metód sú vzájomne porovnané. V prílohe sa nachádzajú tabuľky s výsledkami, funkcie naprogramované v programe Matlab (m-file). Priložené CD obsahuje všetky programové prvky sponínané v tejto práci a vzorky jednotlivých testovaných rečových signálov.

1 REČ – VZNIK, PRENOS, VNÍMANIE

Reč patrí medzi najstaršie komunikačné prvky človeka a patrí medzi najčastejšie používanú formu komunikácie medzi ľuďmi. Hovorená reč sa prenáša komunikačným kanálom vo forme akustických vĺn akustického signálu, ktorého podstatu tvorí vlnenie prostredia v rozsahu počuteľných frekvencií.

Proces vytvárania reči človekom, popísaný v [6, s. 22-32], spočíva v používaní hlasového traktu, ktorý sa dá rozdeliť na dychové ústrojenstvo, hlasové ústrojenstvo a artikulačné ústrojenstvo.

Dychové ústrojenstvo tvorí základný zdroj energie pre reč.

Hlasové ústrojenstvo má na starosti samotný vznik hlasu a je lokalizované v hrtane. Najdôležitejšou časťou hlasového ústrojenstva sú hlasivky. V prípade, že človek mlčí, hlasivky sú úplne otvorené a vzduch pri dýchaní prechádza hrtanom bez odporu. V prípade, že človek začne rozprávať, hlasivky sa uzatvoria, pričom vzduch vychádzajúci z pľúc pôsobí na hlasivky tlakom, tie sa rozkmitajú a začínú sa prudko otvárať a zatvárať. V dôsledku kmitania hlasiviek sa vzduchový prúd zmení s homogénneho na nehomogénny tak, že sa takmer pravidelne strieda kvantum redšieho a hustejšieho vzduchu. Vzniká vzduchová vlna, ktorú vnímame ako zvuk.

Artikulačné ústrojenstvo sa v refazci hlasového traktu nachádza na konci. Umožňuje vytvárať veľké množstvá rôznych zvukov, ktoré pomáhajú charakterizovať jazyk reči. Skladajú sa z artikulačných nástrojov umiestnených v nadhrtanových dutinách (medzi ne patrí hrdelná dutina, nosová dutina a ústna dutina). Zatiaľ, čo artikulačné orgány sa podieľajú na tvorbe reči aktívne, vyššie spomínané dutiny sa na tvorbe reči podieľajú iba pasívne, no majú svoj význam pri tvorbe hlások (základných fonetických jednotiek). Mnohé subjektívne metódy vyberajú testovaný materiál práve podľa toho, či daná hláska vznikla v ústnej dutine, alebo v nosovej dutine [4]. V ústnej dutine totiž vznikajú oblasti so sústredenou akustickou energiou (taktiež sa daným oblastiam hovorí formanty). Akonáhle sa zapojí do tvorby reči aj nosová dutina, tá svojimi antirezonančnými vlastnosťami potláča niektoré frekvencie (vytvára tzv. antiformanty).

Podľa [2] siaha rozsah základných frekvenčných zložiek hlasu od 500 Hz do 3000 Hz

pričom veľmi veľký vplyv na zrozumiteľnosť má pásmo 1000 Hz – 2000 Hz. Ostatné zložky majú skôr vplyv na farbu zvuku.

Celkový proces tvorby reči je veľmi komplikovaný a nie je náplňou tejto práce. Detailnejší popis vzniku a vlastností reči je dobre analyzovaný v [6].

1.1 Kvalita a zrozumiteľnosť

Kvalita, ako aj zrozumiteľnosť sú iba jedným z mnohých atribútov rečového signálu. Nie sú si však rovné. Preto sa používajú iné vyhodnocovacie metódy pre vyhodnotenie kvality rečového signálu a iné pre vyhodnotenie jeho zrozumiteľnosti. Kvalita ako taká je vysoko subjektívna, nakoľko každý z poslucháčov má inú predstavu o tom, čo je „dobrá“ kvalita a čo je „zlá“ kvalita, preto je náročné vyhodnotiť ju presne a spoľahlivo. Kvalita vo svojej podstate má mnoho dimenzií, príliš mnoho, aby sa dali vyčíslieť a obsiahnuť v testovacích algoritmoch. Preto sa pri praktickej aplikácii zameriava iba na niektoré z nich. Pod pojmom zrozumiteľnosť sa rozumie posúdenie toho, čo rečník povedal, tzn. význam a obsah jeho slov.

Na rozdiel od kvality zrozumiteľnosť nepredstavuje pri určovaní až taký veľký problém, nakoľko nie je taká subjektívna a dá sa zistiť pomerne jednoducho. Napríklad prezentovaním hovoreného materiálu (viet, slov, atď.) skupine poslucháčov, ktorá má za úlohu identifikovať hovorené slová. Následne sa zrozumiteľnosť dá určiť spočítaním všetkých slov a spočítaním všetkých správne identifikovaných slov.

Nie je však jednoduché navrhnúť spoľahlivý test, ktorý by vhodne otestoval zrozumiteľnosť. Do hry totiž vstupuje mnoho faktorov. Aby bol test zrozumiteľnosti čo najlepší, mal by podľa [4] obsahovať nasledovné:

1. Dobrá reprezentácia všetkých majoritných hlások: Základné rečové zvuky (fonémy) by mali byť obsiahnuté v každom zozname testových materiálov. Najlepšie by bolo, ak by sa frekvencia výskytu foném v čo najväčšej miere zhodovala s frekvenciou výskytu foném v bežnej konverzácii. Táto podmienka zaručuje, že výsledné skóre bude odrážať komunikačnú situáciu v reálnom svete.
2. Rovnaká náročnosť testových zoznamov: V situácii, kde je potrebné otestovať viacero algoritmov sa používa viacero testových zoznamov, aby si poslucháči

neboli schopní zapamätať testovaný materiál.

3. Kontrola kontextovej informácie: Človek je inteligentný tvor, ktorý pri vnímaní reči uplatňuje aj svoje vedomosti (napr.: sémantika, gramatika, pragmatická atď.), takže dokáže veľmi jednoducho vnímať kontextové informácie, ktoré sú napríklad vo vetách. To robí slová prezentované vo vetách zrozumiteľnejšími ako slová, ktoré sú poslucháčovi prezentované samostatne. Poslucháč teda nemusí rozumieť všetkým slovám vo vete, a pritom stále môže pochopiť význam vety. Kvôli tomu je potrebné pri testoch využívajúcich ako prezentovaný materiál vety kontrolovať množstvo kontextovej informácie. Je mnoho subjektívnych metód pre určenie zrozumiteľnosti rečového signálu. Každá sa ku kontextovej informácii stavia inak. Niektoré danú informáciu kontrolujú, niektoré ju redukujú prezentovaním viacerých slov, z ktorých si poslucháč môže vybrať a niektoré ju odstraňujú úplne prezentovaním slov osamote.

Vzťah medzi zrozumiteľnosťou reči a kvalitou reči nie je plne pochopený a to je čiastočne spôsobené tým, že sa doposiaľ nepodarilo identifikovať akustickú koreláciu kvality a zrozumiteľnosti [4]. hlas prenášaný cez internet – Voice over Internet Protocol (VoIP) je dobrým príkladom pre znázornenie vzťahu medzi kvalitou a zrozumiteľnosťou. Signál prijímaný cez VoIP môže byť vo vysokej kvalite, avšak pri strate paketov nie je zrozumiteľný. Z toho teda vyplýva, že signál môže byť vysokej kvality, ale nemusí byť zrozumiteľný. Platí to aj naopak, signál môže byť dobre zrozumiteľný aj napriek nízkej kvalite. Uvedený príklad poukazuje na to, že kvalita a zrozumiteľnosť sú dva úplne odlišné pojmy, a že je potrebný odlišný prístup pri ich vyhodnocovaní.

1.2 Vplyvy znižujúce kvalitu a zrozumiteľnosť

Každý deň sa všade vôkol nás môžeme stretnúť s rôznymi hlukmi, ktoré sa môžu pri zvukovej nahrávke prejaviť ako nežiadúci šum. Šum je vo všeobecnosti podľa [4] rozdelený na šum, ktorý je v čase konštantný (napríklad šum generovaný ventilátorom) a šum, ktorý sa v čase mení (napríklad šum v reštaurácii). Ďalším zdrojom

šumu môžu byť napríklad kodeky spracúvajúce zvukový signál. Preto sa začali vyvíjať algoritmy, ktoré by tento nežiadúci šum potlačili.

1.3 Algoritmy pre potláčanie šumu

Úlohou týchto algoritmov je takzvané „vylepšiť“ rečový signál, napríklad znížením úrovne šumu. Každý odšumovací algoritmus musí byť navrhnutý tak, aby za každú cenu zachoval zrozumiteľnosť.

Odšumovacie algoritmy sa delia na niekoľko kategórií podľa [4, s. 97-460]. Do skupiny algoritmov *spektrálneho odčítania* patrí jeden z historicky prvých algoritmov na vylepšenie rečového signálu (odstránením šumu). Spomínaný algoritmus spektrálneho odčítania vychádza z jednoduchého princípu, že šum je vlastne iba pripočítaný k pôvodnému signálu, čo znamená, že sa dá zaznamenať v momente, keď nie je prítomný rečový signál a jednoducho odpočítať od zašumeného signálu. Ďalšiu skupinu tvoria algoritmy na štatisticky založenom modeli. To znamená, že sa napríklad spraví séria meraní odpovedajúcich koeficientom fourierovej transformácie zašumeného signálu a hľadá sa nejaký lineárny alebo nelineárny ukazovateľ. Do tejto skupiny patria algoritmy ako *Wienerov algoritmus* a Minimum Mean-Square Error (MMSE) algoritmus. Poslednú skupinu tvoria takzvané *podpriestorové algoritmy* využívajúce väčšinou teóriu lineárnej algebry - na rozdiel od predchádzajúcich. Presnejšie povedané, tieto algoritmy sú založené na princípe, že neskreslený signál môže byť obmedzený na podpriestor zašumeného Euklidovho priestoru. Podrobnejší popis a konkrétne algoritmy sú popísané v [4]. Existuje mnoho algoritmov na zlepšenie rečového signálu. S narastajúcim množstvom algoritmov sa začalo čoraz viac diskutovať o tom, ktorý z algoritmov je lepší.

Nasledujúca časť sa zaoberá spomínanou problematikou. Metódy vyhodnocovania kvality rečového signálu sa delia na subjektívne a objektívne. Medzi objektívne metódy patria intruzívne a neintruzívne metódy. Bližšie vysvetlenie jednotlivých kritérií je v nasledujúcej časti.

1.4 Subjektívne metódy

Subjektívne metódy patria medzi najpresnejšie spôsoby ako spoľahlivo a s čo najväčšou presnosťou určiť kvalitu rečového signálu. Ich podstatou je prezentovanie signálu, ktorého kvalitu chceme zistiť, skupine zväčša trénovaných poslucháčov, ktorí majú za úlohu ohodnotiť daný signál podľa predom stanovenej stupnice. Značnou nevýhodou daných metód je ich náročnosť na trénovaných poslucháčov a náročnosť na čas. Často sú totiž veľmi zdĺhavé, čo ich v konečnom dôsledku robí aj drahými. Výsledky zo subjektívnych metód sa následne môžu použiť pre vyhodnotenie výkonnosti objektívnych metód (Meria sa korelácia výsledkov objektívnych metód s výsledkami subjektívnych metód).

1.4.1 Metóda párového porovnávania

Metóda párového porovnávania, známa tiež ako komparatívny prístup, je podrobne opísaná v [5]. V psychometrii patrí párové porovnávanie medzi teoreticky i technicky najprepracovanejšie metódy. Jeho hlavnou výhodou je to, že pre jeho realizovanie nie je potrebné použiť trénovaných poslucháčov a je možné ho použiť na škálovanie najrôznejších psychologických premenných. Naopak hlavnou nevýhodou párového porovnávania je, najmä pri testovaní väčšieho počtu objektov, značná časová náročnosť úzko súvisiaca s trpezlivosťou pokusných osôb. Príklad pre metódu párového porovnávania je v tabuľke 1.1, kde sa porovnáva 5 podnetov. Jednotlivé podnety sú pokusnej osobe postupne prezentované v dvojiciach. Z každej dvojice sa vyberie lepší, dominantnejší podnet. Tomu podnetu sa udelí vyššia váha (jedna), horšiemu nižšia váha (nula). Celkový počet prezentovaných dvojíc $C_2(n)$ sa dá vyjadriť ako

$$C_2(n) = \binom{n}{2} = \frac{n(n-1)}{2} = 0,5(n^2 - n), \quad (1.1)$$

kde n je počet testovaných podnetov. Neutrálne rozhodnutie je pri použití tejto metódy neprípustné. To znamená, že pokusná osoba sa musí definitívne rozhodnúť pre jeden podnet, ktorý označí ako lepší. Táto technika je v [5] popísaná ako nútená voľba. Preto, ak sa v teste vyskytne dvojica, v ktorej si pokusná osoba nie je istá, môže ju to viesť k tipovaniu. V prípade, že takýchto dvojíc sa v teste nachádza viacero, môže to pokusnú osobu deprimovať, čím sa zhorší jej motivácia a

kvalita posudkov. Vyhodnocovanie výsledkov párového porovnávania a problémy s tým spojené sú podrobnejšie rozobrané v kapitole 2.2.2.

Tab. 1.1: Tabuľka porovnania kvality piatich podnetov metódou párového porovnávania

	P_1	P_2	P_3	P_4	P_5	Σ
P_1	×	1	1	1	0	3
P_2	0	×	0	0	0	0
P_3	0	1	×	1	0	2
P_4	0	1	0	×	0	1
P_5	1	1	1	1	×	4

1.4.2 MOS skóre

Skóre hodnotiace priemerný názor – Mean Opinion Score (MOS) skóre je v súčasnosti najpoužívanějšía priama metóda pre určenie kvality pri subjektívnych testoch. Jej podstata spočíva v tom, že pokusná osoba sa snaží prijímaný podnet ohodnotiť podľa dopredu stanovenej škály. Existuje niekoľko škál MOS, pričom každá je vhodná pre iné účely. V [4] je opísaná škála MOS zobrazená v tab. 1.2. Táto metóda je pomerne náročná na realizáciu. V počiatkovej fáze sa musia pokusné osoby naučiť čo je „výborná“ kvalita, čo je „priemerná“ kvalita a čo je „zlá“ kvalita. Je totiž nevyhnutné, aby sa vo svojich hodnoteniach zhodli, tzn. ak prvá pokusná osoba vníma prezentovaný podnet ako „dobrý“, musí ho tak vnímať aj druhá pokusná osoba. V nasledujúcej fáze sú jednotlivým pokusným osobám v náhodnom poradí prezentované testované podnety k ohodnoteniu. Minimálne požiadavky sú taktiež pomerne vysoké. Minimálny počet laických poslucháčov je 20 a minimálny počet profesionálnych poslucháčov je 10. Okrem toho, že tieto testy sú časovo náročné, sú aj finančne náročné. Výsledné skóre sa získa vypočítaním priemeru získaného skóre pre každý testovaný podnet.

Tab. 1.2: Tabuľka hodnotenia podnetov pomocou skóre MOS

Známka	Kvalita reči	Úroveň skreslenia
5	Výborná	Nepoznatelná
4	Dobrá	Poznatelná, ale nie rušivá
3	Priemerná	Poznatelná a trochu rušivá
2	Slabá	Rušivá, ale nie protivná
1	Zlá	Veľmi rušivá a protivná

1.5 Objektívne metódy

Najlepší spôsob ako spoľahlivo určiť kvalitu reči je použitie subjektívnych metód. Tie však majú mnoho nevýhod. Často vyžadujú veľa času a pre ich zrealizovanie aj trénovaných poslucháčov. Tieto dôvody viedli mnohých k myšlienke nahradiť subjektívne metódy objektívnymi. Teda metódami, ktoré nevyžadujú k vyhodnoteniu kvality rečového signálu poslucháčov. Ideálna metóda by mala byť schopná určiť kvalitu rečového signálu aj bez prístupu k pôvodnému signálu. Ideálna vyhodnocovacia metóda by teda mala obsahovať nízkoúrovňové spracovanie, napríklad psychoakustika (maskovanie zvukov, frekvencia, harmonické tóny atď.) a vysokoúrovňové spracovanie ako sémantika (význam slova), pragmatika (používanie jazyka v praxi), prozódia (eufonická kvalita hlások, kvantita, slabičnosť, prízvuk, dôraz, predel, melódia, výška tónu a pauza) a lingvistika. Ak by existovala objektívna metóda, ktorá by dokázala vyhodnocovať rečový signál podľa vyššie uvedených požiadaviek, jej výsledok by sa s vysokou presnosťou približoval výsledku klasických subjektívnych testov využívajúcich profesionálnych poslucháčov. V oblasti vývoja objektívnych metód bol spravený veľký pokrok. Väčšina dnešných objektívnych metód sa stále limituje na potrebu použitia pôvodného signálu. Navzdory faktu, že väčšina metód používa nízkoúrovňové spracovanie, ich výsledky sa veľmi dobre porovnávajú s výsledkami subjektívnych testov (napr. skóre MOS). Podstata objektívnych metód spočíva v porovnávaní testovaného signálu s referenčným signálom. Tieto signály sa medzi sebou neporovnávajú priamo, ale najprv sa rozdelia na rámce, obvykle

s dĺžkou 10 – 30 **ms**!. Následne sa počíta skreslenie jednotlivých testovaných rámcov voči referenčným rámcem. To sa môže udiť v časovej doméne alebo vo frekvenčnej doméne.

1.5.1 Intruzívne metódy

Intruzívne metódy sú metódy, ktoré pre určenie kvality signálu vyžadujú prístup k pôvodnému, originálnemu signálu. Z [4, k. 10.5] sú v tejto práci rozobrané tri objektívne intruzívne metódy a jedna prezentovaná v [8].

Meranie segmentálnej miery SNR v časovej doméne

Aby malo dané meranie zmysel, musia byť odstránené akékoľvek fázové chyby alebo časové posuny medzi pôvodným signálom a spracovávaným signálom. Signály sa následne rozdelia na rámce, pričom dĺžka jedného rámca sa volí v rozmedzí 15 ms – 20 ms. Potom segmentálna miera pomer signálu a šumu – Signal to Noise Ratio (SNR) sa vypočíta ako

$$\text{SNR}_{\text{seg}} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \frac{\sum_{n=Nm}^{Nm+N-1} x^2(n)}{\sum_{n=Nm}^{Nm+N-1} (x(n) - \hat{x}(n))^2}, \quad (1.2)$$

kde $x(n)$ predstavuje originálny signál, $\hat{x}(n)$ hodnotený signál, N je dĺžka rámca a M je počet rámcov. Pri použití tejto metódy však nastáva jeden problém. Týka sa tichých pasáží v rečovom signáli, ktoré sú v bežnej reči dosť časté. Energia signálu v týchto miestach je veľmi malá, čo bude mať za následok vysoké záporné hodnoty segmentálnej miery SNR, ktoré skreslia celkový výsledok merania. Je viacero spôsobov ako sa danému problému vyhnúť. Jedným z nich je spôsob, kedy sa zo sumy z rovnice (1.2) vylúčia tiché pasáže porovnaním merania krátkodobej energie oproti prahu alebo zaokrúhlením hodnôt SNR_{seg} na nízku hodnotu. Ďalším možným riešením je limitovať rozsah hodnôt SNR_{seg} na interval -10 dB až 35 dB, čím sa dá vyhnúť nutnosti detekovať tiché pasáže v rečovom signáli.

Zrejme najúčinnější a najjednoduchší spôsob ako tomu predísť je prezentovaný v [7], kde je posunutá logaritmická funkcia o hodnotu jedna.

$$\text{SNR}_{\text{segR}} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \left(1 + \frac{\sum_{n=Nm}^{Nm+N-1} x^2(n)}{\sum_{n=Nm}^{Nm+N-1} (x(n) - \hat{x}(n))^2} \right), \quad (1.3)$$

Tým sa dosiahlo, že najnižšia hodnota, ktorú je možné dostať je nula namiesto pôvodného mínus nekonečna, čím sa dá vyhnúť skresleniu merania a nie je viac potrebné používať detekciu tichých pasáží.

Meranie segmentálnej miery SNR vo frekvenčnej doméne

Segmentálna miera SNR môže byť určená ako v časovej, tak aj vo frekvenčnej doméne. Vo frekvenčnej doméne sa teda vypočíta ako

$$\text{fwSNR}_{\text{seg}} = \frac{10}{M} \sum_{m=0}^{M-1} \frac{\sum_{j=1}^K B_j \log_{10} \left[\frac{F^2(m,j)}{(F(m,j) - \hat{F}(m,j))^2} \right]}{\sum_{m=0}^{M-1} B_j}, \quad (1.4)$$

kde B_j predstavuje váhu pripadajúcu j-tému frekvenčnému pásmu, K je počet pásiem, M je celkový počet rámcov v signáli, $F(m, j)$ je amplitúda m-tého rámcu originálneho signálu pri j-tej frekvencii pásma banky filtrov a $\hat{F}(m, j)$ je amplitúda spracovávaného signálu toho istého rámca pri tej istej frekvencii. V tabuľke 1.3 sú zobrazené váhy priradené jednotlivým frekvenciám váhových filtrov. Frekvencie v spomínanej tabuľke boli získané za predpokladu použitia tretino-oktávových filtrov a váhy priradené jednotlivým frekvenciám boli interpolované tak, aby poskytl ekvivalentnú váhu pre rozteč filtrov kritických pásiem. Uvedené váhy nie sú fixne stanovené, to znamená, že rôznym frekvenciám môže byť priradená váha podľa potreby. To udeľuje danej metóde flexibilitu, ktorá je značnou výhodou oproti (1.2) a (1.3).

Tab. 1.3: Tabuľka váh pre centrálnu kmitočty filtrov

Číslo pásma	Centrálna frekvencia (Hz)	Váha	Číslo pásma	Centrálna frekvencia (Hz)	Váha
1	50	0,003	14	1148	0,032
2	120	0,003	15	1288	0,034
3	190	0,003	16	1442	0,035
4	260	0,007	17	1610	0,037
5	330	0,010	18	1794	0,036
6	400	0,016	19	1993	0,036
7	470	0,016	20	2221	0,033
8	540	0,017	21	2446	0,030
9	617	0,017	22	2701	0,029
10	703	0,022	23	2978	0,027
11	798	0,027	24	3276	0,026
12	904	0,028	25	3592	0,026
13	1020	0,030			

Normalizácia rámcov

Metóda určovania kvality reči pomocou normalizácie rámcov využíva odhad LSE. Signál $x(t)$ predstavuje originálny, neskrútený signál. Signál $y(t)$ predstavuje signál skreslený, alebo inak upravený. Pre určenie kvality signálu $y(t)$ je potrebné odmerať vzdialenosť $y(t)$ od $x(t)$. Podobne, ako v predchádzajúcich prípadoch, je najprv potrebné signály normalizovať a rozdeliť na rámce. Jednotlivé rámce sa následne prevedú do frekvenčnej oblasti pomocou Hanningovho okna a FFT. Pre prípad, aby tiché rušivé pasáže nezasahovali do výpočtov sú v [8] oba signály logaritmicke transformované a orezané podľa

$$\tilde{x}_i = \max(10 \cdot \log_{10}(x_i), -40). \quad (1.5)$$

LSE sa následne vypočíta ako

$$e_t = \frac{1}{u-l+1} \sum_{i=l}^u |\tilde{y}_i - \tilde{x}_i|. \quad (1.6)$$

Existuje aj iný spôsob ako interpretovať (1.6). Tento spôsob navyše zahŕňa kroky, kde sa najprv zistí priemerná hodnota oboch spektier a následne sa normalizuje výstupné spektrum

$$e_c = \frac{1}{u-l+1} \sum_{i=l}^u \tilde{y}_i - \frac{1}{u-l+1} \sum_{i=l}^u \tilde{x}_i, \quad \hat{y} = \tilde{y}_i - e_c. \quad (1.7)$$

Celkové LSE e_t (1.6) je teda rozdelené na dve časti. Hrubý komponent e_c (1.7) a jemný komponent e_f (1.8).

$$e_f = \frac{1}{u-l+1} \sum_{i=l}^u |\hat{y}_i - \tilde{x}_i|. \quad (1.8)$$

Pri vyhodnocovaní sa uvažuje s dvojpásmovou implementáciou (1.6) a (1.6). Prvé pásmo je v rozmedí od jednosmernej zložky do približne 2 kHz ($l = 2, u = 64$), čo je vo výsledku označené ako e_{tl} a e_{cl} . Druhé pásmo sa rozprestiera od 2 kHz do Nyquistovej frekvencie ($l = 65, u = 128$), čo sa značí ako e_{th} a e_{ch} . Aritmetické priemery jednotlivých výsledkov sú označené \bar{e}_{tl} , \bar{e}_{th} , \bar{e}_{cl} , \bar{e}_{ch} a \bar{e}_f . Podľa [8] sa následne pri použití parametrov z tab. 1.4 vypočíta kvalita ako

$$\hat{q} = \left[1 + \exp \left(w_0 + \sum_{i=1}^M w_i \cdot p_i \right) \right], \quad (1.9)$$

pričom výsledok $R_{str} \in \langle 0; 1 \rangle$ a je kladne korelovaný k výsledkom MOS testov pozostávajúcich zo 17 sád testov obsahujúcich cez 12 000 nahrávok za použitia 320 kodekov).

Tab. 1.4: Tabuľka parametrov a váh využitých pri výpočte R_{str}

i	p_i	w_i
0		-4,29
1	\bar{e}_{cl}	0,150
2	\bar{e}_{ch}	0,0243
3	\bar{e}_f	0,449

1.5.2 Neintruzívne metódy

Doteraz rozoberané objektívne metódy vyžadovali prístup k pôvodnému signálu (kvôli odmeraniu „vzdialenosti“ spracovávaného signálu od pôvodného). Mnohokrát sa však vyžaduje odmeranie kvality signálu, pričom pôvodný signál nie je k dispozícii, čo robí intruzívne metódy nepoužiteľnými. Dobrým príkladom je VoIP, kde sa požaduje pokiaľ možno kontinuálne monitorovanie kvality signálu počas hovoru. V [4] je spomenutých niekoľko typov neintruzívnych metód. Medzi nimi viaceré metódy, ktoré sa opierajú o umelo vyrvorený referenčný signál odvodený zo špeciálnych tabuliek a číselníkov, alebo metóda, ktorá najprv vyberie z testovaného signálu parametre hlasového traktu a vyhodnotí, či dané zvuky mohli byť vygenerované človekom a podľa veľkosti jednotlivých parametrov určí kvalitu. Táto oblasť určovania kvality reči je zatiaľ pomerne neprebádaná vzhľadom k jej možnostiam. Navyše je veľmi obtiažné zohnať nejakú literatúru, ktorá by popisovala neintruzívne metódy podrobnejšie.

2 PRAKTICKÁ ČASŤ PRÁCE

2.1 Tvorba databázy nahrávok

Príprava testovaného materiálu pozostávala z niekoľkých krokov:

1. Prvým krokom bolo nahrávanie, ktoré prebiehalo v mŕtvej komore. Pre tieto účely bolo vybraných 6 rečníkov (tri ženy a traja muži vo veku 18 – 22 rokov). Každý rečník predčítal 6 – 7 krátkych viet v slede za sebou, ktoré boli pomocou štyroch mikrofónov (vzdialených od rečníka 1 meter) zaznamenávané do programu *CuBase 4*.
2. Druhým krokom bolo opäť nahrávanie avšak v tomto prípade bol každému rečníkovi do slúchadiel pustený rôzny šum. Ako prvý šum *Babble* (hluk z jedálne počas obeda), následne šum *Factory* (hluk vytvorený výrobnou linkou vo fabrike), *White noise* (biely šum), a ul. Zubri (hluk nahraný z ulice Zubří). Pre tento druh experimentu boli použité referenčné štúdiové slúchadlá *Samson SR850* s poloootvorenou konštrukciou. Výsledkom sú nahrávky, v ktorých je možné sledovať Lombardov efekt (najviac sa prejavuje pri šume ul. Zubri nakoľko tento šum má premenlivú dynamiku, ktorá sa mení s prechádzajúcimi motorovými vozidlami).
3. Tretí krok spočíval v normalizovaní nahrávok na 0 dB, ich následnom rozstrihaní na jednotlivé vety a k nim pripadajúce šumy. Potom boli nahrávky zašumené príslušným šumom s odstupom 0 dB a 6 dB. Pre normalizovanie, strihanie a zašumovanie nahrávok bol použitý software *Audacity 1.3 Beta*.
4. V štvrtom kroku boli vybrané 2 nahrávky, ktoré boli odšumené pomocou odšumovacích algoritmov. Pre tieto účely boli použité: *spektrálne mapovanie* a *spektrálne odčítanie* pri nastavení parametra p na hodnoty $p = 0,2$, $p = 0,5$ a $p = 0,8$ (parameter p má vplyv na odhad veľkosti šumu). Výsledné nahrávky sú zostavené do jednoduchej databázy (túto databázu je možné nájsť na priloženom multimediálnom disku).

Nahrávky, ktoré sú pri tomto experimente použité boli zaznamenané kvalitným kondenzátorovým mikrofónom *Rode NT3* určeným na snímanie hlasu. V tab. 2.1 sú

zobrazené údaje o jednotlivých nahrávkach. Dĺžka jednej nahrávky je v rozmedzí 6 – 8 sekúnd. Predčítaný text znie: „Nepoznám kľúč k úspechu, viem však, že kľúčom k neúspechu je snaha zapáčiť sa každému“ a „Naša pamäť sa nedá použiť ani ako šnúrka do topánok, taká je krátka!“.

Tab. 2.1: Zoznam použitých nahrávok

Nahrávka	Rečník	Šum (SNR 6dB)	Odšumovací algoritmus
A	Žena 22	žiadny	žiadny
B	Žena 22	B/F/WN/UZ	spektrálne mapovanie
C	Žena 22	B/F/WN/UZ	spektrálne odčítanie $p = 0,2$
D	Žena 22	B/F/WN/UZ	spektrálne odčítanie $p = 0,5$
E	Žena 22	B/F/WN/UZ	spektrálne odčítanie $p = 0,8$
F	Žena 22	B/F/WN/UZ	žiadny
G	Muž 22	žiadny	žiadny
H	Muž 22	B/F/WN/UZ	spektrálne mapovanie
I	Muž 22	B/F/WN/UZ	spektrálne odčítanie $p = 0,2$
J	Muž 22	B/F/WN/UZ	spektrálne odčítanie $p = 0,5$
K	Muž 22	B/F/WN/UZ	spektrálne odčítanie $p = 0,8$
L	Muž 22	B/F/WN/UZ	žiadny

2.2 Psychoakustický experiment

2.2.1 Návrh

Pri návrhu psychoakustického experimentu sa vychádzalo z [5]. Hlavným cieľom bolo predísť faktorom, ktoré by mohli mať za následok skreslenie údajov. Medzi ne patrí únava pokusných osôb, stupeň zacvičenia pokusných osôb a Fechnerove chyby miesta a času.

Únave pokusných osôb je možné predísť vhodnou voľbou dĺžky experimentu. Pre posluchové testy nesmie podľa [4] dĺžka experimentu prekročiť 20 minút, aby ne-

došlo k zníženiu pozornosti a k zhoršeniu úsudku pokusných osôb. Preto bol test koncipovaný tak, aby aj pre pokusné osoby s dlhším rozhodovaním netrval viac ako spomínaných 20 minút. Pod pojmom Fechnerovej chyby miesta sa rozumie fakt, že jeden prvok z dvojice je nejako umiestnený (vľavo, vpravo, hore, dole). V prípade, ak sa pokusná osoba nevie rozhodnúť, ktorý z dvoch prezentovaných podnetov je lepší, môže tak urobiť na základe polohy podnetu (na obrazovke, či odpovedovom hárku), to znamená, že napríklad, podnet umiestnený na prvej pozícii mu bude sympatickejší. Fechnerova chyba času môže nastať, ak sa pri párovom porovnávaní pokusnej osobe prezentujú rovnaké podnety v krátkom časovom slede za sebou. Všetkým spomínaným chybám sa dá predísť vhodným usporiadaním testovaných podnetov. Bol teda zvolený sled dvojíc zobrazený v tab. 2.2. Rossov plán usporiadania dvojíc opísaný v [5] nebolo možné použiť, nakoľko je použiteľný iba pre nepárny počet testovaných podnetov. V tomto experimente je použitých 6 podnetov (počet je zvolený vzhľadom k časovej náročnosti – ak by sa použilo 7 podnetov, celý test by sa značne predĺžil), takže poradie dvojíc bolo vymyslené tak, aby sa žiadny podnet z dvojice neobjavil hneď v nasledujúcej dvojici, aby jednoznačné dvojice neboli zoskupené, a aby neboli zoskupené ani nejednoznačné dvojice. Takéto zoskupenie nejednoznačných dvojíc by mohlo u pokusnej osoby vyvolať pocit, že je to veľmi ťažké a mohlo by viesť k jej deprimovaniu natoľko, že by to malo vplyv na presnosť jej úsudkov. Z toho dôvodu je v pravidelných intervaloch vždy jedna jednoznačná dvojica pre posilnenie sebavedomia pokusnej osoby.

Tab. 2.2: Poradie dvojíc nahrávok pri párovom porovnávaní

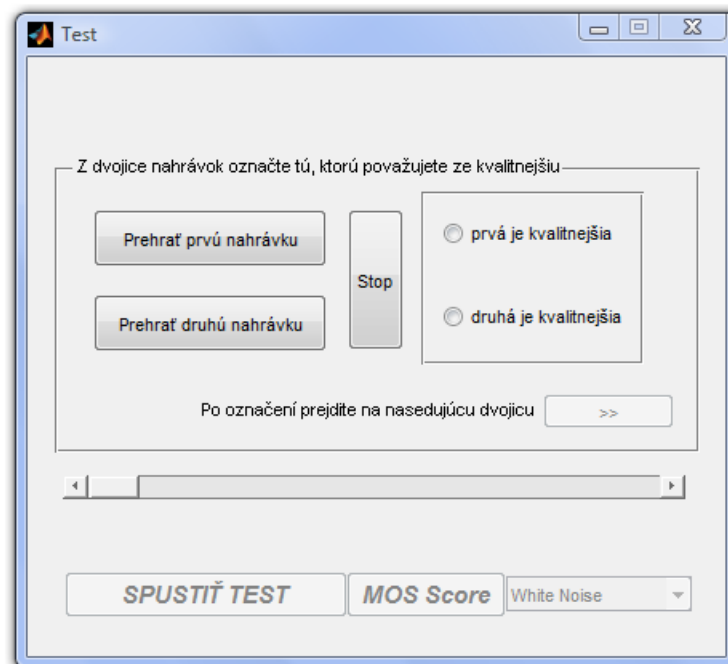
Šum	Poradie dvojíc
Babble	1 2; 3 4; 5 6; 1 3; 2 6; 3 5; 1 4; 2 3; 4 5; 1 6; 2 5; 3 6; 2 4; 1 5; 4 6
Factory	4 6; 1 5; 2 4; 3 6; 2 5; 1 6; 4 5; 2 3; 1 4; 3 5; 2 6; 1 3; 5 6; 3 4; 1 2
White Noise	2 1; 4 3; 6 5; 3 1; 6 2; 5 3; 4 1; 3 2; 5 4; 6 1; 5 2; 6 3; 4 2; 5 1; 6 4
Ul. Zubri	6 4; 5 1; 4 2; 6 3; 5 2; 6 1; 5 4; 3 2; 4 1; 5 3; 6 2; 3 1; 6 5; 4 3; 2 1

Ako je spomínané v sekcii 1.4.2, MOS skóre je najpoužívanejšou metódou na hodnotenie kvality rečového signálu. Napriek tomu je veľmi obtiažné nájsť podrobnejšie informácie o priebehu testu (hlavne čo sa týka fázy učenia pokusných osôb). Vedľajším cieľom tohto experimentu bolo vyskúšať hodnotenie pomocou MOS skóre aj napriek nesplneniu podmienok spomínaných v 1.4.2.

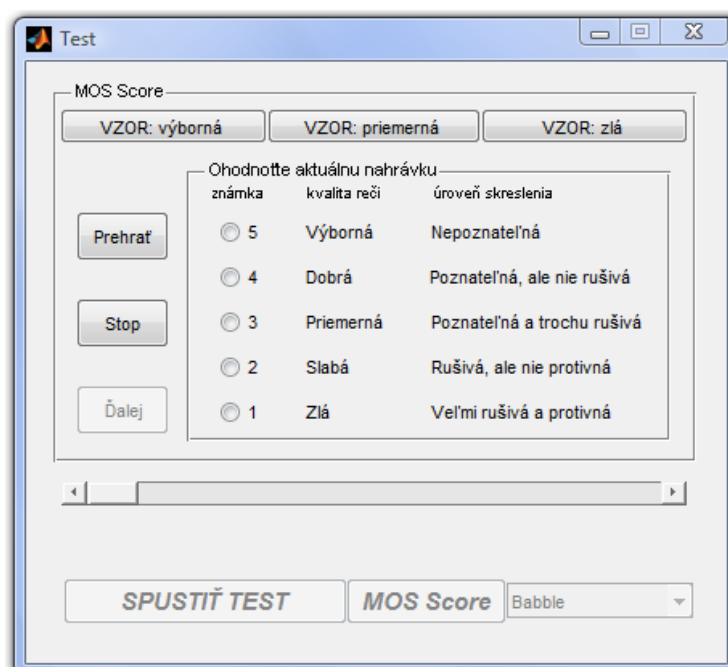
Pre zjednodušenie celého experimentu bola v programe Matlab vyvinutá GUI aplikácia s jednoduchým dotazníkom zobrazeným na obrázku Obr. 1, prostredím pre párové porovnávanie – Obr. 2 a prostredím pre hodnotenie nahrávok pomocou MOS skóre – Obr. 3. Výstup údajov z tejto aplikácie je vyriešený pomocou textového súboru. Dotazník má informatívny charakter a má za úlohu trochu zatraktívniť test. Výsledky tohto dotazníka budú spracovávané tak, aby odpovede jednotlivých pokusných osôb zostali anonymné.

The screenshot shows a MATLAB GUI window titled "Test". Inside the window, there is a section titled "Údaje o poslucháčovi" (Listener Data) which contains several input fields and radio buttons. The fields are: "Meno" (Name) with the value "Peter", "Priezvisko" (Surname) with the value "Tuleja", and "Vek" (Age) with the value "22". Below these is a "Pohlavie" (Gender) section with two radio buttons: "Muž" (Male) which is selected, and "Žena" (Female). There are two questions with "Áno" (Yes) and "Nie" (No) radio button options: "Myslíte si, že máte hudobný sluch?" (Do you think you have a musical ear?) with "Áno" selected, and "Ovládáte hru na nejaký hudobný nástroj?" (Do you play any musical instrument?) with "Áno" selected. Below these is a text input field for "Aký?" (Which?) with the value "Bicie, akordeón" (Drums, accordion). At the bottom of the form is a question "Zúčastnili ste sa už niekedy posluchových testov?" (Have you ever participated in listening tests?) with "Áno" (Yes) and "Nie" (No) radio buttons, where "Nie" is selected. Below the form is a label "Správnosť údajov potvrdte stlačením:" (Confirm the correctness of the data by clicking:) followed by an "OK" button. At the very bottom of the GUI are three buttons: "SPUŠTIŤ TEST" (Run Test), "MOS Score", and a dropdown menu labeled "--vyberte test--" (select test).

Obr. 1: Dotazník



Obr. 2: Párové porovnávanie



Obr. 3: MOS skóre

2.2.2 Výsledky

Experimentu sa zúčastnilo celkovo 14 poslucháčov, pričom finálnej verzii testu sa ich z časových dôvodov zúčastnilo iba 10 (poslucháči boli vyberaní nenáhodným príležitostným výberom, to znamená, že sú to prevažne študenti, ktorí mali čas a chuť sa experimentu zúčastniť). Poslucháči vyplňovali test jednotlivo. Posluch bol uskutočnený vždy v tom istom prostredí za použitia referenčných štúdiových slúchadiel *Samson SR850*. Celý experiment bol rozdelený na dva pokusy. Prvý pokus obsahoval nahrávky A – F 2.1. Druhý pokus sa konal s časovým odstupom dvoch dní a obsahoval nahrávky G – L 2.1. Každému poslucháčovi boli prezentované nahrávky obsahujúce 3 rôzne šumy. V tabuľke 2.3 je znázornený systém výberu šumov pre jednotlivých poslucháčov, ktorý sa cyklicky opakuje po každých štyroch poslucháčoch. Pri výbere variantov sa dbalo aj na to, aby niektorá z pokusných osôb nedostala v oboch pokusoch ten istý variant. Každá pokusná osoba teda mala možnosť vypočuť si a ohodnotiť nahrávky obsahujúce všetky použité šumy.

Tab. 2.3: Systém výberu šumov pre jednotlivých poslucháčov

	varianta 1	varianta 2	varianta 3	varianta 4
Babble	✓	✗	✓	✓
Factory	✓	✓	✗	✓
White noise	✓	✓	✓	✗
ul. Zubri	✗	✓	✓	✓

Párové porovnávanie

Pri párovom porovnávaní môže u jednotlivých pokusných osôb dochádzať k javu, ktorý je v [5] opísaný ako nekonzistentnosť. Povedzme, že pokusná osoba označí $P_1 > P_2$ a $P_2 > P_3$. Z toho sa dá logicky odvodiť, že $P_1 > P_3$. Ak však pokusná osoba označí $P_1 < P_3$ vznikne kruhová trojica a takáto odpoveď sa považuje za nekonzistentnú. Čím viac kruhových trojíc sa v odpovediach jednej osoby vyskytne,

tým je jej výsledok menej konzistentný. V tomto smere je veľmi dobrým ukazovateľom Kendallov koeficient konzistencie spomenutý v [5]. Môže nadobúdať hodnoty v intervale $K \in \langle 0; 1 \rangle$, kde $k = 0$ značí žiadnu konzistenciu a naopak $k = 1$ maximálnu konzistenciu. Jeho výpočet pre párny počet testovaných podnetov je uvedený vo vzťahu (2.1), kde R_j znamená súčet preferencií v j -tom riadku matice a R_{str} je aritmetický priemer súčtu preferencií a dá sa zapísať ako $(\sum R_j)/n$.

$$k = \frac{12 \sum (R_j - R_{\text{str}})^2 - 3n}{n^3 - 4n} \quad (2.1)$$

Pre jednoduchosť je nižšie uvedený konkrétny vzorový výpočet pre jednu preferenčnú maticu.

Tab. 2.4: Matica preferencií jednotlivca pre nahrávky obsahujúce šum z ul. Zubří

	A	B	C	D	E	F	R_j	$(R_j - R_{\text{str}})^2$
A	×	1	1	1	1	1	5	6,25
B	0	×	1	1	1	1	4	2,25
C	0	0	×	1	1	0	2	0,25
D	0	0	0	×	1	1	2	0,25
E	0	0	0	0	×	0	0	6,25
F	0	0	1	0	1	×	2	0,25
\sum	0	1	3	3	5	3	15	15,5

Pre uvedený príklad teda platí, že $n = 6$, $R_{\text{str}} = 15/6$, $\sum (R_j - R_{\text{str}})^2 = 15,5$. Po dosadení do vzťahu (2.1) dostaneme $k = 0,875$.

Keďže odvodzovanie kruhových trojíc z odpovedí pokusnej osoby je veľmi pracné (obzvlášť pri vyššom počte testovaných podnetov), v [5] je počet kruhových trojíc definovaný ako

$$d = \frac{n(n^2 - 4)(1 - k)}{24}. \quad (2.2)$$

Následne môžeme dosadiť vypočítanú hodnotu $k = 0,875$ do vzťahu (2.2), čím dostaneme výsledok $d = 1$. To znamená, že v uvedenej matici preferencií sa nachádza jedna kruhová trojica. Ak je cieľom dosiahnuť štatisticky významný výsledok na hladine významnosti 1% tak podľa [5] je maximálny možný počet kruhových trojíc pre

počet podnetov $n = 6$ stanovený na $d = 0$. Vypočítaný výsledok nespĺňa danú podmienku, takže musí byť vyradený z vyhodnocovania. V prípade tohto experimentu je však cieľom dosiahnuť štatisticky významný výsledok na hladine významnosti 5%, čo znamená, že podľa [5] je prípustný počet kruhových trojíc stanovený na $d = 2$. Tejto podmienke daná preferenčná matica vyhovuje, takže môže byť zaradená do vyhodnocovania. Tento postup sa zopakuje pri každej preferenčnej matici.

Po uskutočnení vyššie uvedených výpočtov pre každú maticu a po vyradení matíc, ktoré nespĺňajú podmienku $d \leq 2$, je možné jednotlivé matice zlúčiť do spoločnej matice preferencií. Tak, ako bolo potrebné zistiť konzistenciu jednotlivca, tak je potrebné zistiť konzistenciu aj v rámci skupiny. Pre tento účel poslúži Kendallov koeficient zhody spomínaný v [5]. Pri maximálnej zhode v skupine je jeho veľkosť $u_{\max} = 1$, pri maximálnej nezhode je pre párny počet n jeho veľkosť vyjadrená ako $u_{\min} = -1/(m - 1)$, čo v tomto prípade, pre $m = 7$ platných preferenčných matíc, je $u_{\min} = -0,16\bar{6}$.

Tab. 2.5: Matica preferencií skupiny pre nahrávky obsahujúce šum z ul. Zubří

	A	B	C	D	E	F	R_j
A	×	7	7	7	7	7	35
B	0	×	2	2	6	2	12
C	0	5	×	7	6	4	22
D	0	5	0	×	6	3	14
E	0	1	1	1	×	0	3
F	0	5	3	4	7	×	19
Σ	0	23	13	21	32	16	105

V prílohe je možné nájsť skupinovú preferenčnú maticu z výsledkami pre prvý aj druhý pokus.

Kendallov koeficient zhody sa podľa [5] vypočíta ako

$$u = \frac{8 \sum}{m(m-1)n(n-1)} - 1. \quad (2.3)$$

kde m je počet pokusných osôb, n je počet testovaných podnetov a \sum sa vypočíta ako

$$\sum = \sum_{\substack{j,k=1 \\ j \neq k}}^n C_2(p_{jk}) = \sum_{\substack{j,k=1 \\ j \neq k}}^n \binom{p_{jk}}{2} = 0,5 \sum_{\substack{j,k=1 \\ j \neq k}}^n (p_{jk}^2 - p_{jk}). \quad (2.4)$$

Tabuľka 2.6 obsahuje vypočítané hodnoty $C_2(p_{jk})$.

Tab. 2.6: Matica hodnôt $C_2(p_{jk})$ pre nahrávky obsahujúce šum z ul. Zubří

	A	B	C	D	E	F	R_j
A	×	21	21	21	21	21	105
B	0	×	1	1	15	1	18
C	0	10	×	21	15	6	52
D	0	10	0	×	15	3	28
E	0	0	0	0	×	0	0
F	0	10	3	6	21	×	40
\sum	0	51	25	49	87	31	243

Hľadanú hodnotu koeficientu zhody dostaneme dosadením do vzťahu (2.3). Pre $m = 7$, $n = 6$ a $\sum = 243$ sa $u = 0,5429$. Ide o pomerne vysokú hodnotu koeficientu zhody. Pre kontrolu môže byť otestovaná štatistická významnosť dosiahnutého výsledku pomocou χ^2 testu podľa vzťahu

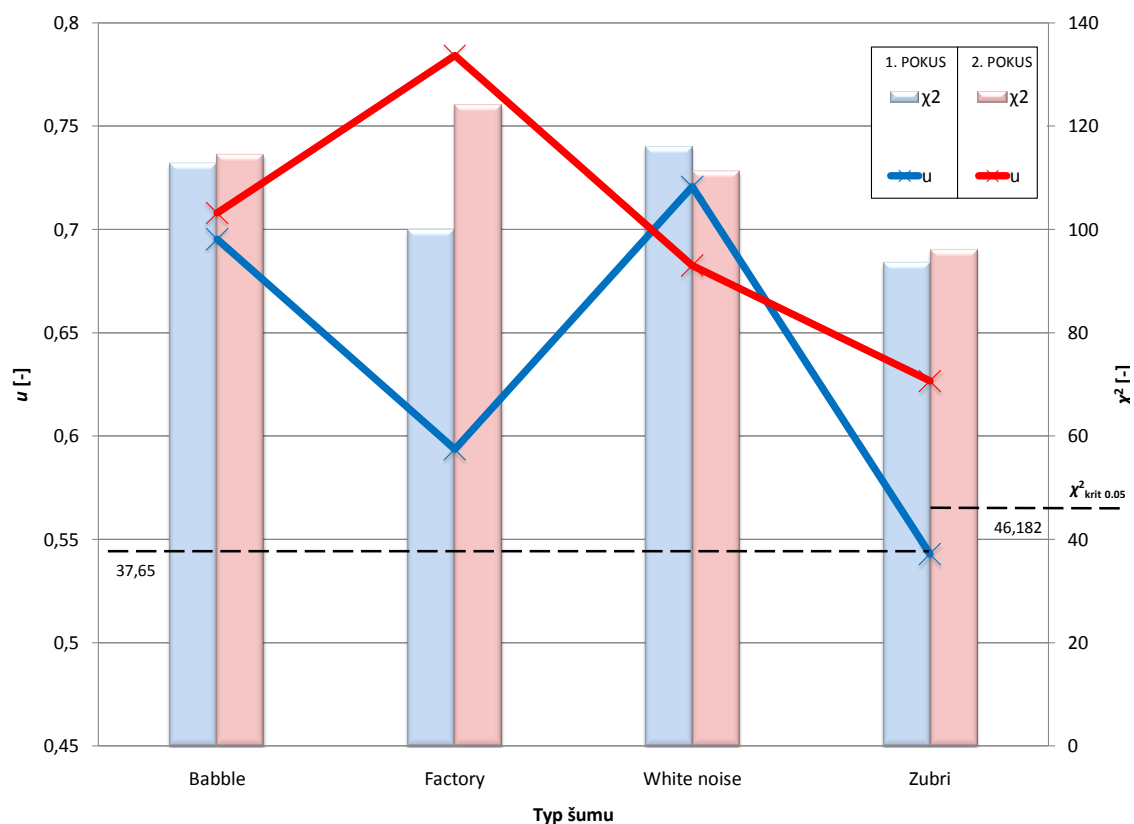
$$\chi^2 = \frac{4}{m-2} \left[\sum - \frac{1}{2} \binom{n}{2} \binom{m}{2} \frac{m-3}{m-2} \right], \quad (2.5)$$

pre

$$s_v = \binom{n}{2} \frac{m(m-1)}{(m-2)^2} \quad (2.6)$$

stupňov voľnosti. Pre uvedený príklad to po dosadení do (2.6) prakticky znamená $s_v = 25,2$ a po vypočítaní (2.5) $\chi^2 = 93,6$. Z tabuľky kritických hodnôt χ^2 [1] je možné pre $s_v = 25$ zistiť $\chi_{\text{krit } 0,05}^2 = 37,65$. Keďže vypočítaná hodnota χ^2 je takmer 2,5 krát väčšia ako kritická hodnota zistená z tabuľky, môže byť hodnota $u = 0,5429$ považovaná za štatisticky významnú hodnotu. V grafe 2.1 je možné vidieť výsledné koeficienty zhody pre jednotlivé šumy pri oboch pokusoch. Štatistická významnosť je

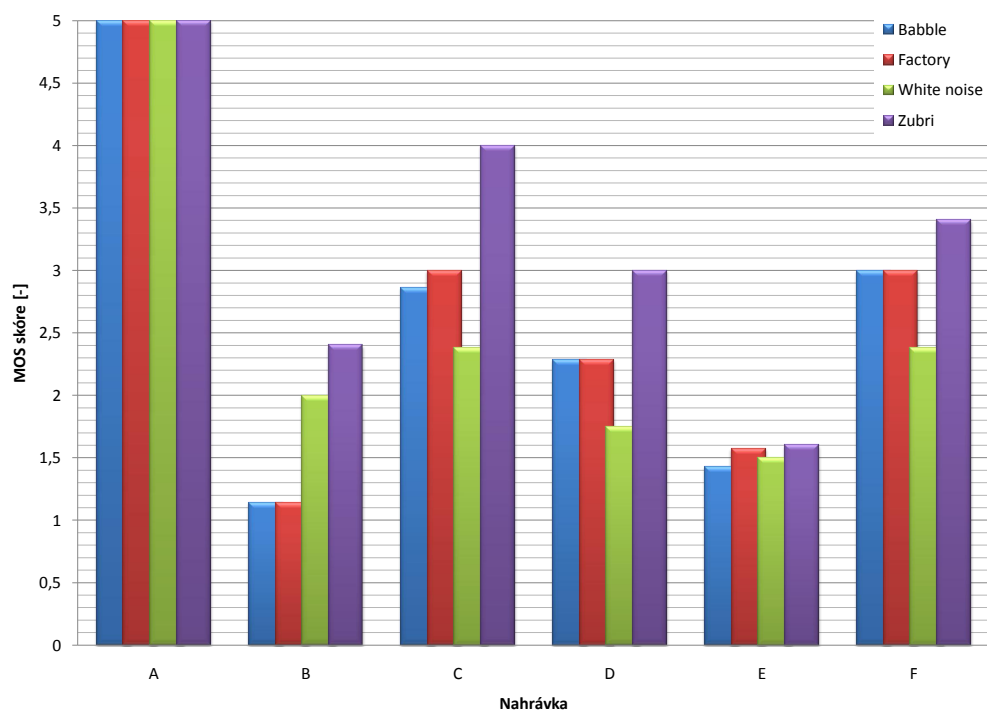
v danom grafe znázornená pomocou vypočítaných hodnôt χ^2 a z tabuľky zistených hodnôt $\chi^2_{\text{krit } 0,05}$.



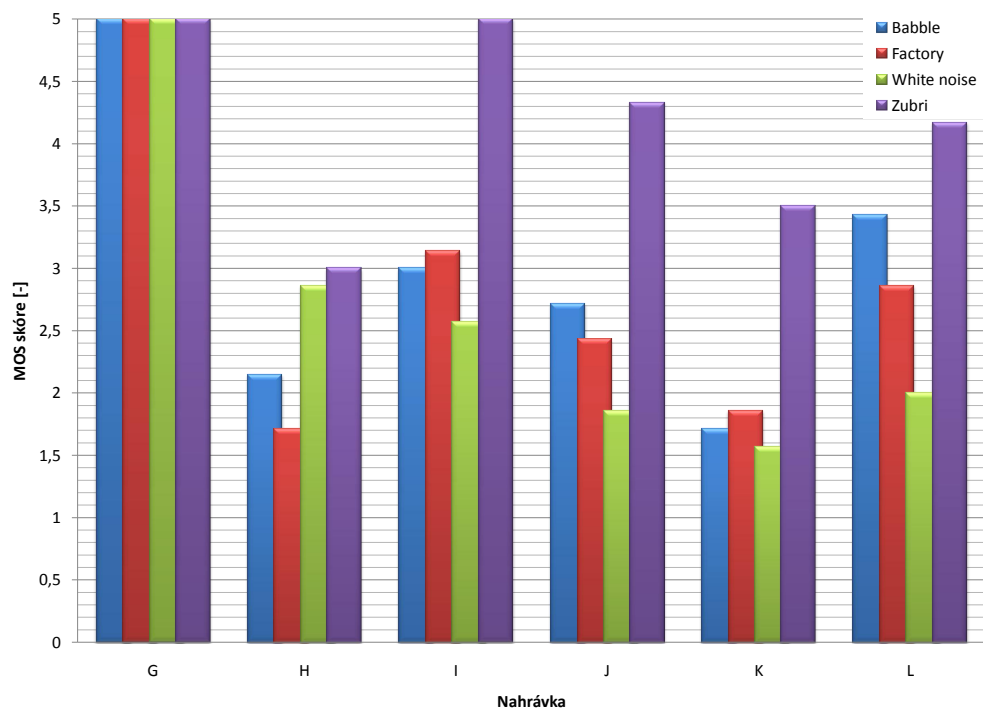
Graf 2.1: Graf koeficientov zhody a štatistickej významnosti

MOS skóre

Výsledné MOS skóre sa počíta ako aritmetický priemer všetkých známok udelených danému signálu. Výsledky MOS skóre pre prvý pokus sú znázornené v grafe 2.2, výsledky pre druhý pokus sú v grafe 2.3.



Graf 2.2: Graf výsledkov MOS skóre pre 1. pokus



Graf 2.3: Graf výsledkov MOS skóre pre 2. pokus

2.3 Objektívne metódy

2.3.1 Implementácia

Pre implementáciu metód opísaných v sekcii 1.5.1 bol použitý program Matlab R2009b. Základnú časť programu tvorí súbor *Main.m*, kde sa nachádzajú všetky uživateľom nastavované prvky (signály, ktoré majú byť porovnané, dĺžka rámca, váhy pre jednotlivé frekvencie a nakoniec údaje pre výpočet korelácie). M-file *Main.m* využíva pri výpočtoch niekoľko vytvorených funkcií:

- *Framework.m* má za úlohu odstrániť akékoľvek časové posuny medzi danými signálmi, normalizovať ich amplitúdu a upraviť dĺžku tak, aby bolo možné rozdeliť signál na rámce s dopredu definovanou dĺžkou.
- *SegSNR.m* počíta SNR v časovej oblasti podľa (1.2) a (1.3).
- *FreqSNR.m* počíta SNR vo frekvenčnej oblasti podľa (1.4).
- *framenorm.m* počíta kvalitu postupom opísaným v 1.5.1
- *correlation.m* počíta koreláciu medzi subjektívnou a objektívnou metódou podľa 2.7.

Bližší popis jednotlivých funkcií je možné nájsť v prílohe.

2.3.2 Výsledky

V tejto sekcii sú tabuľky obsahujúce výsledky objektívnych metód pre všetky signály.

Tab. 2.7: Výsledky SNRseg

	A	B	C	D	E	F
Babble	60,008681	-8,815085	-1,977543	-1,371802	-0,909183	-5,452546
Factory	60,008681	-8,634321	3,702714	3,488639	2,968609	-2,004971
White noise	60,008681	0,295736	3,537467	3,774132	4,122150	-2,994513
ul. Zubri	60,008655	-11,072184	2,670760	2,943833	2,992051	-1,562810
	G	H	I	J	K	L
Babble	60,008646	-5,393403	0,301419	0,502132	0,661220	-5,241984
Factory	60,008680	-2,442618	3,761216	3,668208	3,487984	-0,900194
White noise	60,008681	-0,486400	3,898183	4,429640	5,031321	-2,516265
ul. Zubri	60,008680	-8,356411	15,414574	15,459759	15,164881	12,021686

Tab. 2.8: Výsledky SNRsegR

	A	B	C	D	E	F
Babble	60,008685	1,021315	3,256801	3,380526	3,276927	1,919947
Factory	60,008686	0,815439	6,291103	6,227799	5,952866	2,892158
White noise	60,008685	4,636547	7,462492	7,263313	6,825806	3,391294
ul. Zubri	60,008659	0,720644	6,881273	7,103820	7,160047	4,086083
	G	H	I	J	K	L
Babble	60,008650	2,401293	5,435986	5,395706	5,239579	2,571933
Factory	60,008684	2,878038	6,667174	6,508080	6,212972	3,803772
White noise	60,008686	3,342431	7,215422	7,229229	6,949650	3,431098
ul. Zubri	60,008684	0,765219	16,776510	16,727595	16,451678	14,318352

Tab. 2.9: Výsledky fwSNRseg

	A	B	C	D	E	F
Babble	60,002663	-3,689873	7,910854	7,950492	7,647289	4,172812
Factory	60,005343	-5,646858	8,725525	8,650962	8,443931	4,278933
White noise	60,008682	3,101568	7,995637	8,202663	8,228757	2,971043
ul. Zubri	60,003056	-5,929483	12,639588	12,712816	12,608251	10,038637
	G	H	I	J	K	L
Babble	59,947815	-0,248846	6,961642	7,015752	6,811602	2,741913
Factory	60,008682	1,282541	8,853060	8,917435	8,806182	5,201063
White noise	60,008682	1,866393	8,447497	8,509888	8,405858	3,048574
ul. Zubri	60,008682	-6,363366	20,954541	20,935063	20,705612	20,618890

Tab. 2.10: Výsledky framenorm

	A	B	C	D	E	F
Babble	0,986477	0,817157	0,848973	0,867780	0,895275	0,804270
Factory	0,986477	0,882847	0,874952	0,887187	0,912091	0,829345
White noise	0,986474	0,903041	0,797540	0,828594	0,880734	0,708750
ul. Zubri	0,986462	0,934568	0,921412	0,926184	0,935168	0,898256
	G	H	I	J	K	L
Babble	0,986476	0,812252	0,818582	0,831783	0,855666	0,769475
Factory	0,986477	0,910260	0,891350	0,900625	0,918497	0,851058
White noise	0,986479	0,934737	0,829883	0,853559	0,893410	0,751055
ul. Zubri	0,986478	0,963856	0,953938	0,958261	0,965918	0,935577

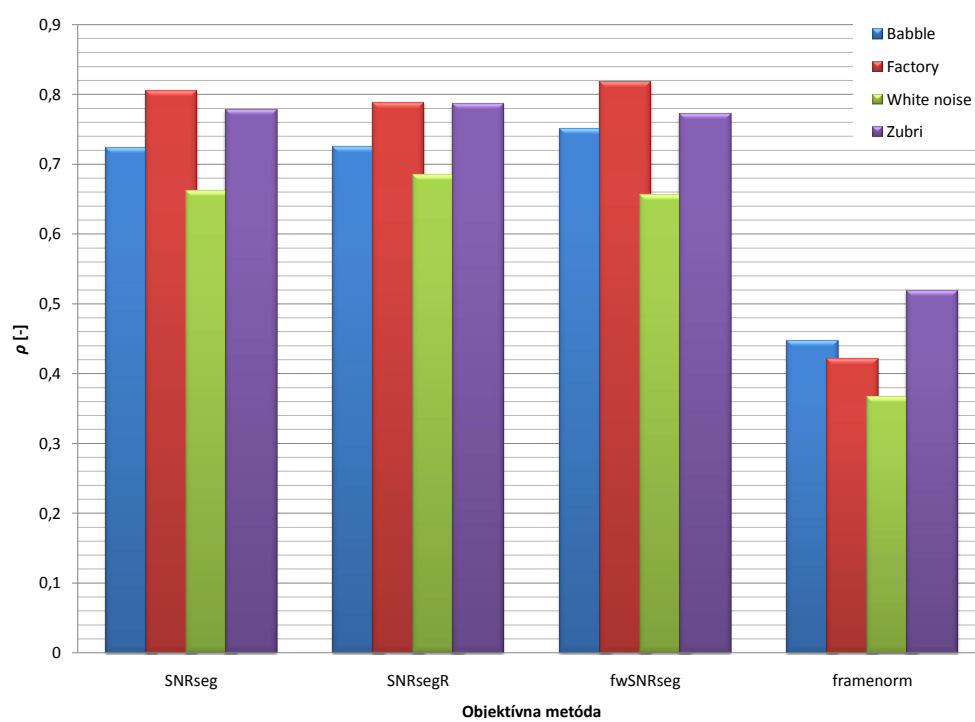
2.4 Vyhodnotenie objektívnych metód

Nakoľko je možné z tabuliek v sekcii 2.3.2 odvodiť, ktorý signál je lepší a ktorý je horší – dá sa jednoducho určiť ich poradie od najlepšieho po najhorší, ale takýto postup nie je práve najvhodnejším riešením. Zoradením nahrávok podľa poradia sa totiž odstráni akákoľvek informácia o vzájomnej vzdialenosti.

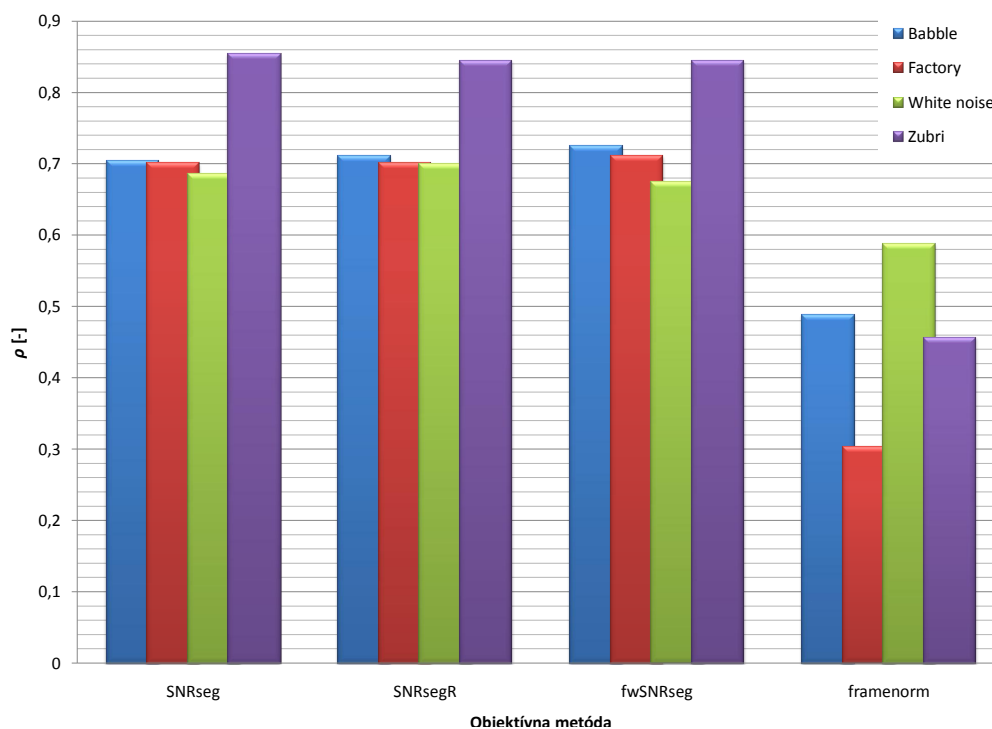
Pre stanovenie výsledkov s oveľa vyššiou výpovednou hodnotou je potrebné určiť koreláciu výsledkov subjektívnej metódy s jednotlivými objektívnymi metódami. Preto boli výsledky metód spracované podľa [3], kde sa korelácia subjektívnej a objektívnej metódy určí ako

$$\rho = \frac{\sum_d (S_d - \bar{S}_d)(O_d - \bar{O}_d)}{\sqrt{\sum_d (S_d - \bar{S}_d)^2} \sqrt{\sum_d (O_d - \bar{O}_d)^2}}, \quad (2.7)$$

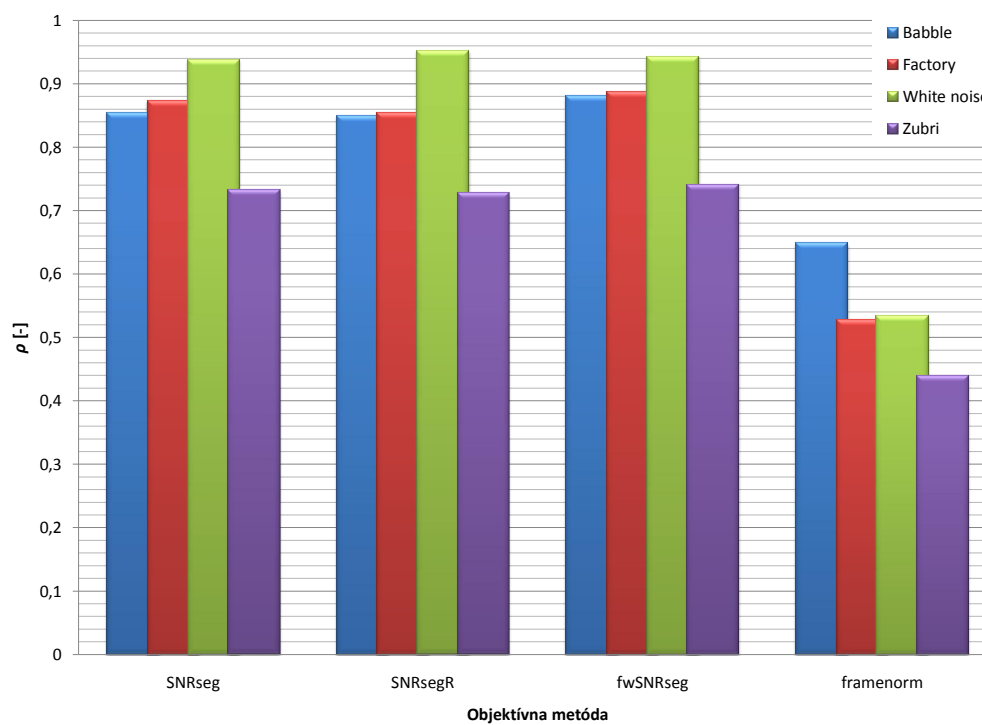
pričom S_d predstavuje subjektívnu metódu, O_d objektívnu metódu a \bar{S}_d , \bar{O}_d sú ich stredné hodnoty. Po dosadení dosiahnutých výsledkov jednotlivých metód do rovnice (2.7) boli dosiahnuté výsledky uvedené v nasledujúcich grafoch. Graf 2.4 znázorňuje koreláciu párového porovnávania s jednotlivými objektívnymi metódami pre 1. pokus a graf 2.5 znázorňuje koreláciu párového porovnávania s jednotlivými objektívnymi metódami pre 2. pokus. Na grafe 2.6 je možné vidieť koreláciu objektívnych metód a MOS skóre pre prvý pokus a na grafe 2.7 koreláciu objektívnych metód a MOS skóre pre druhý pokus.



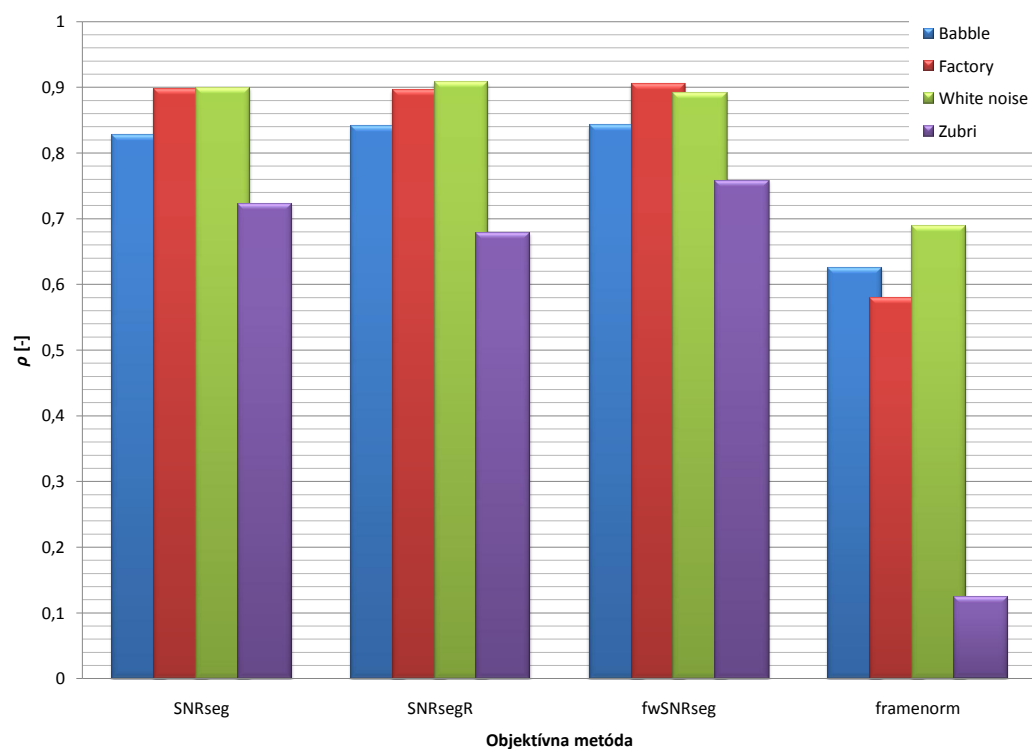
Graf 2.4: Graf korelácie výsledkov párového porovnávania a objektívnych metód pre 1. pokus



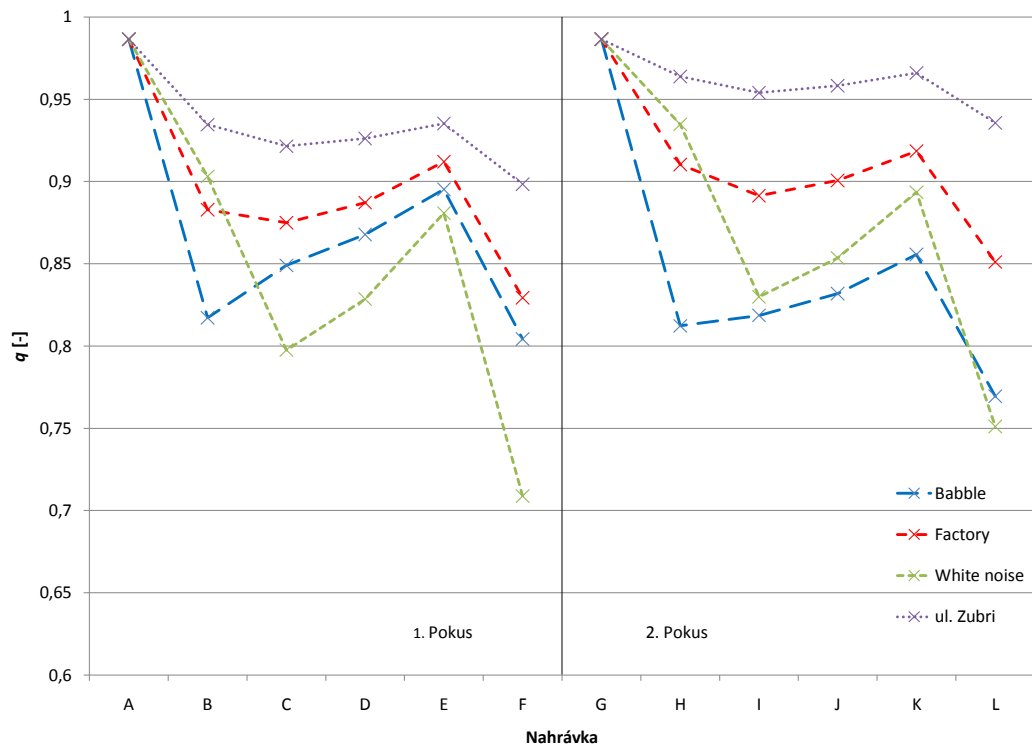
Graf 2.5: Graf korelácie výsledkov párového porovnávania a objektívnych metód pre 2. pokus



Graf 2.6: Graf korelácie výsledkov MOS skóre a objektívnych metód pre 1. pokus



Graf 2.7: Graf korelácie výsledkov MOS skóre a objektívnych metód pre 2. pokus



Graf 2.8: Graf korelácie normalizácie rámcov a MOS skóre z [8] pre 1. a 2. pokus

ZÁVER

Cieľom tejto práce je navrhnúť a vyskúšať rámcový systém pre určenie kvality rečového signálu. V teoretickej časti sú opísané základné náležitosti, ktoré sa týkajú reči, jej spracovania a vyhodnocovania. Zo subjektívnych metód je podrobne popísaná metóda párového porovnávania a metóda ohodnocovania pomocou MOS skóre. Z objektívnych metód sú rozobrané dve metódy merajúce mieru SNR v časovej doméne, jedna metóda merajúca mieru SNR vo frekvenčnej doméne a metóda normalizácie rámcov využívajúca odhad LSE. Praktická časť sa zaoberá návrhom samotného rámcového systému. V sekcii 2.1 je popísaný spôsob akým boli nahrávky vytvorené, spracované a zašumené. Tabuľka 2.1 obsahuje zoznam a vlastnosti použitých nahrávok.

Návrh subjektívneho testu mal niekoľko fáz kým sa dostal do podoby aká je opísaná v tejto práci. Návrh sa zameriaval hlavne na metódu párového porovnávania. Boli odskúšané rôzne počty testovaných podnetov, rôzne poradie testovaných podnetov, rôzne dĺžky testovaných podnetov. Napriek tomu, že bol prevedený iba nenáhodný príležitostný výber pokusných osôb z úzkej skupiny ľudí (študenti prevažne rovnakého veku), dá sa povedať, že výsledok stačí na overenie funkčnosti navrhnutého rámcového systému. V sekcii 2.2.2 sa podľa rovnice (2.1) počítaná koeficient zhody pokusnej osoby. Nízka priemerná hodnota koeficientu zhody môže znamenať zle navrhnutý test obsahujúci mnoho nejednoznačných dvojíc. Priemerná hodnota koeficientu zhody $\bar{k} = 1$ hovorí taktiež o zle navrhnutom teste pozostávajúcom zväčša z jednoznačných dvojíc. Priemerná hodnota koeficientu zhody pri prvom pokuse je $\bar{k} = 0,9598$ a pri druhom pokuse je $\bar{k} = 0,9722$. Jedná sa o pomerne vysoké koeficienty zhody no je vidieť, že testované dvojice neboli úplne jednoznačné no neboli ani úplne nejednoznačné.

Ako je možné vidieť v sekcii 2.4 sú dosiahnuté pomerne vysoké hodnoty korelácie subjektívnych a objektívnych metód. Napriek tomu, že boli pri počítaní korelácie objektívnych metód a MOS skóre dosiahnuté vyššie hodnoty, nedá sa povedať, že je metóda hodnotenia pomocou MOS skóre presnejšia (bola navrhovaná iba okrajovo). Párové porovnávanie je v tomto experimente oveľa prepracovanejšie, čo znamená, že

jeho výsledky sa dajú považovať za presnejšie.

Vysoké hodnoty korelácie subjektívnych a objektívnych metód potvrdzujú funkčnosť navrhnutého rámcového systému pre vyhodnocovanie kvality rečového signálu. Aby však tento systém dosahoval spoľahlivé výsledky je potrebné zmeniť niekoľko vecí. Rozšíriť databázu nahrávok (k tomu sa prihliadalo už pri nahrávaní, preto bola nahraná rozsiahlejšia databáza obsahujúca 6 rečníkov, 13 rôznych predčítaných textov v dvoch jazykoch s dĺžkou 5 – 25 sekúnd pri štyroch šumoch, čo je približne 200 nahrávok. Keďže bolo nahrávanie snímané pomocou štyroch rôznych mikrofónoch celkovo je to cez 800 zvukových nahrávok, ktoré sú v neupravenej a nepostrihanej forme priložené na DVD), zmeniť spôsob výberu a počet pokusných osôb.

LITERATÚRA

- [1] ANDĚL, Jiří. *Statistické metody*. Praha : Matfyzpress, 2007. 300 s. ISBN 80-7378-003-8.
- [2] EHRENFRIED, Martin. Improving the intelligibility of SSB transmissions : What is important when processing speech?. Technical Feature [online]. February 2009, no. 2, [cit. 2011-06-01]. Dostupný z WWW: <<http://g8jnj.webs.com/SSB%20intelligibility%20part%201.pdf>>.
- [3] HU, Yi; LOIZOU, Philipos C. Evaluation of Objective Quality Measures for Speech Enhancement. IEEE Transactions on Audio, Speech, and Language Processing [online]. Január 2008, vol. 16, no. 1, [cit. 2010-12-12]. Dostupný z WWW: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=&rep=rep1&type=pdf>>.
- [4] LOIZOU, Philipos C. *Speech Enhancement : Theory and Practice*. Boca Raton : CRC Press, 2007. 608 s. ISBN 978-0-8493-5032-0.
- [5] MELKA, Alois. *Základy experimentální psychoakustiky*. Vyd. 1. Praha : AMU, 2005. 327 s. ISBN 80-7331-043-0.
- [6] PSUTKA, Josef, et al. *Mluvíme s počítačem česky*. Vyd. 1. Praha : Academia, 2006. 752 s. ISBN 80-200-1309-1.
- [7] RICHARDS, D. L. *Speech transmission performance of PCM systems*. Electron. Lett. 1965, 1, 40-41. Dostupný z WWW: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4205397>>.
- [8] VORAN, S. Advances in objective estimation of perceived speech quality. *Proceedings of the 1999 IEEE speech coding workshop* [online]. 23.06.1999, [cit. 2011-06-01]. Dostupný z WWW: <http://www.its.bldrdoc.gov/pub/audio/pubs_talks/scw99.pdf>.

ZOZNAM SYMBOLOV, VELIČÍN A SKRATIEK

B_j	váha pre j-té frekvenčné pásmo
dB	decibel
d	počet kruhových trojíc
$C_2(n)$	počet neopakujúcich sa dvojíc
e_c	Hrubá zložka LSE
e_f	Jemná zložka LSE
e_t	celková zložka LSE
$F(m, j)$	amplitúda m-tého rámca originálneho signálu pri j-tej frekvencii
$\hat{F}(m, j)$	amplitúda m-tého rámca spracovávaného signálu pri j-tej frekvencii
fwSNRseg	segmentálny pomer signálu a šumu vo frekvenčnej doméne
Hz	hertz
k	Kendallov koeficient konzistencie
K	počet frekvenčných pásiem
LSE	Log-Spectral Error
M	počet rámcov
MMSE	Minimum Mean-Square Error
MOS	Skóre hodnotiace priemerný názor – Mean Opinion Score
N	dĺžka rámca
O_d	výsledok objektívnej metódy
\bar{O}_d	stredná hodnota výsledku objektívnej metódy
p	parameter ovplyvňujúci veľkosť odhadu šumu

p_{jk}	preferencia ležiaca v j -tom riadku a k -tom stĺpci
\hat{q}	Odhad kvality kladne kolerovaný k MOS
R_j	súčet preferencií
R_{str}	aritmetický priemer súčtu preferencií
S_d	výsledok subjektívnej metódy
\bar{S}_d	stredná hodnota výsledku subjektívnej metódy
SNR	pomer signálu a šumu – Signal to Noise Ratio
SNRseg	segmentálny pomer signálu a šumu v časovej doméne
u	Kendallov koeficient zhody
VoIP	hlas prenášaný cez internet – Voice over Internet Protocol
$x(n)$	originálny signál
$\hat{x}(n)$	hodnotený signál
$x(t)$	originálny signál
$y(t)$	hodnotený signál
χ^2	kritická hodnota chí-kvadrát
$\chi^2_{\text{krit } 0,05}$	kritická hodnota chí-kvadrát pre hladinu významnosti 5%
ρ	korelácia

A VÝSLEDKY PÁROVÉHO POROVNÁVANIA

Tab. A.1: Babble, pokus 1.

	A	B	C	D	E	F	R_j
A	×	7	7	7	7	7	35
B	0	×	0	1	4	1	6
C	0	7	×	6	7	1	21
D	0	6	1	×	6	1	14
E	0	3	0	1	×	0	4
F	0	6	6	6	7	×	25
Σ	0	29	14	21	31	10	105

Tab. A.2: Factory, pokus 1.

	A	B	C	D	E	F	R_j
A	×	7	7	7	7	7	35
B	0	×	1	0	4	1	6
C	0	6	×	6	6	4	22
D	0	7	1	×	5	0	13
E	0	3	1	2	×	1	7
F	0	6	3	7	6	×	22
Σ	0	29	13	22	28	13	105

Tab. A.3: White noise, pokus 1.

	A	B	C	D	E	F	R_j
A	×	7	7	7	7	7	35
B	0	×	3	7	7	5	22
C	0	4	×	7	7	5	23
D	0	0	0	×	6	1	7
E	0	0	1	0	×	0	1
F	0	2	2	6	7	×	17
Σ	0	13	12	28	34	18	105

Tab. A.4: ul. Zubri, pokus 1.

	A	B	C	D	E	F	R_j
A	×	7	7	7	7	7	35
B	0	×	2	2	6	2	12
C	0	5	×	7	6	4	22
D	0	5	0	×	6	3	14
E	0	1	1	1	×	0	3
F	0	5	3	4	7	×	19
Σ	0	23	13	21	32	16	105

Tab. A.5: Babble, pokus 2.

	G	H	I	J	K	L	R_j
G	×	7	7	7	7	7	35
H	0	×	0	2	3	0	5
I	0	7	×	6	7	4	24
J	0	5	1	×	7	1	14
K	0	4	0	0	×	0	4
L	0	7	3	6	7	×	23
Σ	0	30	11	21	31	12	105

Tab. A.6: Factory, pokus 2.

	G	H	I	J	K	L	R_j
G	×	7	7	7	7	7	35
H	0	×	0	1	4	1	6
I	0	7	×	7	7	5	26
J	0	6	0	×	7	0	13
K	0	3	0	0	×	0	3
L	0	6	2	7	7	×	22
Σ	0	29	9	22	32	13	105

Tab. A.7: White noise, pokus 2.

	G	H	I	J	K	L	R_j
G	×	7	7	7	7	7	35
H	0	×	5	7	7	6	25
I	0	2	×	5	7	6	20
J	0	0	2	×	7	4	13
K	0	3	0	0	×	1	1
L	0	6	2	7	7	×	11
Σ	0	10	15	22	34	24	105

Tab. A.8: Zubri, pokus 2.

	G	H	I	J	K	L	R_j
G	×	5	5	5	5	5	25
H	0	×	1	1	1	1	4
I	0	4	×	5	5	5	19
J	0	4	0	×	4	4	12
K	0	4	0	1	×	1	6
L	0	4	0	1	4	×	9
Σ	0	21	6	13	19	16	75

B ZDROJOVÉ KÓDY

B.1 Main.m

```
%% ===== TIME DOMAIN ===== %%
clc; clear all;

%
%
[x,FS]=wavread('TEST.SIGNALS\Babble\MUZ_22_3_normalized__clean_sig_');x=x.*1.001;
%
[y,FS]=wavread('TEST.SIGNALS\Babble\MUZ_22_3_normalized__clean_sig_');
%[y,FS]=wavread('TEST.SIGNALS\Babble\MUZ_22_3_norm_bab_SNR6dB_SpecMa');
%[y,FS]=wavread('TEST.SIGNALS\Babble\MUZ_22_3_norm_bab_SNR6dB_SSp0_2');
%[y,FS]=wavread('TEST.SIGNALS\Babble\MUZ_22_3_norm_bab_SNR6dB_SSp0_5');
%[y,FS]=wavread('TEST.SIGNALS\Babble\MUZ_22_3_norm_bab_SNR6dB_SSp0_8');
%[y,FS]=wavread('TEST.SIGNALS\Babble\MUZ_22_3_norm_bab_SNR6dB_babble');
[q] = framenorm(x, y,FS);
FrameLength=20e-03;
[l, rx, ry, delay]= Framework(x,y,FS, FrameLength);
[SNRseg,SNRsegR]=SegSNR(rx, ry);
%% ===== FREQUENCY DOMAIN ===== %%
CF = [0 50 120 190 260 330 400 470 540 617 703 798 904 1020 1148 1288 1442 1610
      1794 1993 2221 2446 2701 2978 3276 3697 44100];
CFW = [3 3 3 7 10 16 16 17 17 22 27 28 30 32 34 35 37 36 36 33 30 29 27 26 26
      0.00001]./1000;
[fwSNRseg]=FreqSNR(CF, CFW, rx, ry, FS, 1);
%% ===== CORRELATION ===== %%
Sd = [35 5 24 14 4 23]; %Bab
Od = [60.008646 -5.393403 0.301419 0.502132 0.661220 -5.241984];
[ cor ] = corelation(Sd, Od);
```

Užívateľom nastavované parametre:

premenná x obsahuje vstupný signál

premenná y obsahuje hodnotený signál

premenná $FrameLength$ obsahuje dĺžku rámca v sekundách

premenná Sd výsledky subjektívnej metódy

premenná Od výsledky objektívnej metódy

Pokročilejšie nastavenia:

CF Centrálné frekvencie

CFW Váhy priradené odpovedajúcim frekvenciám.

B.2 Framework.m

```
% =====  
%> @brief Brief description of the Framework function  
%>  
%> This function prepare input signals for further processing mainly by  
%> dividing signals into frames.  
%>  
%> @param x Contains input signal x.  
%> @param y Contains input signal y.  
%> @param FS Contains sampling frequency  
%> @param FrameLength Contains length of frame in milliseconds  
%> @retval l Return number of samples per frame.  
%> @retval rx Return signal x reshaped into matrix where one row is one  
%> frame.  
%> @retval ry Same as rx but this contains signal y.  
%> @retval delay Returns delay between x and y.  
% =====  
  
function [l,rx,ry,delay] = Framework( x,y,FS, FrameLength )  
fsx=FS;x=x';  
fsy=FS;y=y';  
l=floor(FS*FrameLength); %pocet vzoriek v jednom ramci  
delay=finddelay(x,y); %zistenie oneskorenia  
y=circshift(y,[1 -delay]); %oprava oneskorenia  
%x=x.*(1/abs(max(x))); %normalizacia x  
%y=y.*(1/abs(max(y)));  
if (length(x)>length(y)) %vyrovnanie dlzky signalov  
    y(1,length(x))=0;  
else x(1,length(y))=0;  
end  
a=rem(length(x),l); %zistenie potrebnej dlzky  
a=(1-(a/l))*l;  
a=length(x)+a;  
x(1,a)=0; %doplnenie nulami  
y(1,a)=0;  
rx=reshape(x,l,[]);rx=rx'; %rozdelenie na ramce  
ry=reshape(y,l,[]);ry=ry';  
end
```

B.3 segSNR.m

```
%
%=====
%> @brief Brief description of the segSNR function
%> This function provides segmental SNR calculations in time domain.
%> Following equations are used to compute SNR in time domain.
%> \latexonly
%> \begin{equation}\label{SNRseg} \mathrm{SNRseg}=\frac{10}{M}\sum_{m=0}^{M-1}\backslash
\mathrm{log}_{-10}\frac{\sum_{n=Nm}^{Nm+N-1}x^2(n)}{\sum_{n=Nm}^{Nm+N-1}\mathrm{Big}(x(n)
-\widehat{x}(n)\mathrm{Big})^2},
%> \end{equation}
%> \begin{equation}\label{SNRsegR} \mathrm{SNRseg}_{-R}=\frac{10}{M}\sum_{m=0}^{M-1}\backslash
\mathrm{log}_{-10}\left(1+\frac{\sum_{n=Nm}^{Nm+N-1}x^2(n)}{\sum_{n=Nm}^{Nm+N-1}\mathrm{Big}(x(n)-\widehat{x}(n)\mathrm{Big})^2}\right),
%> \end{equation}
%> \endlatexonly
%> @param rx Contains signal x reshaped into matrix where one row is one frame.
%> @param ry Same as rx but this contains signal y
%> @retval SNRseg
%> @retval SNRsegR
%=====

function [ SNRseg,SNRsegR ] = SegSNR( rx , ry )
dim=size( rx );
b(1,dim(1))=0;
c=b;d=b;e=b;
for i=1:dim(1);
    b(i)=sum ((rx(i,:)).^2);
    c(i)=sum ((rx(i,:)-ry(i,:)).^2);
    d(i)=log10(b/c);
    if d(i)<-10^6;
        d(i)=-10^6;
    else d(i)=d(i);
    end
    e(i)=log10(1+(b/c));
end
SNRseg=(10/dim(1))*sum(d);
SNRsegR=(10/dim(1))*sum(e);
fid = fopen('SNRseg.txt','a+');
fseek(fid,0,'eof');
fprintf(fid,'%f',SNRseg);
fclose(fid);
fid = fopen('SNRsegR.txt','a+');
fseek(fid,0,'eof');
fprintf(fid,'%f',SNRsegR);
fclose(fid);
end
```

B.4 FreqSNR.m

```
%
%=====
%> @brief Brief description of the fwSNRseg function
%>
%> This function provides segmental SNR calculations in frequency domain.
%> Following equation is used to compute SNR in frequency domain.
%> \latexonly
%> \begin{equation}
%> \mathrm{fwSNRseg}=\frac{10}{M}\sum_{m=0}^{M-1}\frac{\sum_{j=1}^KB_j\mathrm{log}_{-10}\Big[\frac{F^2(m,j)}{\big(F(m,j)-\widehat{F}(m,j)\big)^2}\Big]}{\sum_{m=0}^{M-1}B_j},
%> \end{equation}
%> \endlatexonly
%> @param CF Contains center filter frequencies (Hz)
%> @param CFW Contains corresponding articulation index weights used for
%> computing weighted spectral distance measure.
%> @param rx Contains signal x reshaped into matrix where one row is one
%> frame.
%> @param ry Same as rx but this contains signal y
%> @param FS Contains sampling frequency
%> @param l Contains number of samples per frame.
%> @retval fwSNRseg
%=====

function [ fwSNRseg ] = FreqSNR( CF, CFW, rx, ry, FS, l )
CFF=1:(FS/l):FS;
ftrx=abs(fft(rx));
ftry=abs(fft(ry));
dim=size(rx);
seg(dim(1))=0;
for m=1:dim(1);
    wrx=0;
    for j=1:length(CF)-1;
        w=(CFF>=CF(j) & CFF<CF(j+1));
        wrx(j)=(CFW(j)*log10(((w.*ftrx(m,:)).^2)/(w.*ftrx(m,:)-w.*ftry(m,:)).^2));
    end
    seg(m)=sum(wrx)/sum(CFW);
end
fwSNRseg=(10/dim(1))*sum(seg);
fid = fopen('fwSNRseg.txt','a+');
fseek(fid,0,'eof');
fprintf(fid,'%f ',fwSNRseg);
fclose(fid);
end
```

B.5 corelation.m

```
% =====
%> @brief Brief description of the corelation function
%>
%> This function compute correlation between results of subjective and
%> objective measures.
%> \latexonly
%> \begin{equation}\label{corr}
%> \mathrm{\rho}=\frac{\sum_{d}(S_{d}-\bar{S}_{d})(O_{d}-\bar{O}_{d})}{\sqrt{\sum_{d}(S_{d}-\bar{S}_{d})^2}\sqrt{\sum_{d}(O_{d}-\bar{O}_{d})^2}},
%> \end{equation}
%> \endlatexonly
%> @param Sd Contains results of subjective measure in a row vector.
%> @param Od Contains results of objective measure in a row vector.
%> @retval cor Returns correlation between Sd and Od.
% =====

function [ cor ] = corelation( Sd, Od )
SSd=mean(Sd);
SOd=mean(Od);
for i=1:length(Sd);
    a(i)=(Sd(i)-SSd)*(Od(i)-SOd);
    b(i)=(Sd(i)-SSd)^2;
    c(i)=(Od(i)-SOd)^2;
end
cor=sum(a)/(sqrt(sum(b))*sqrt(sum(c)));
fid = fopen('cor.txt', 'a+');
fseek(fid,0,'eof');
fprintf(fid, '%f ',cor);
fclose(fid);
end
```

B.6 framenorm.m

```
function [q] = framenorm( x, y, FS )
fsx=FS;x=x';
fsy=FS;y=y';
FrameLength=32e-03;
overlap=50;
l=floor(FS*FrameLength/2); %pocet vzoriek v jednom ramci
delay=finddelay(x,y);
if delay~=0;
    y=circshift(y,[1 -delay]);
end
%x=x.*(1/max(x));
%y=y.*(1/max(y));
if (length(x)>length(y))
    y(1,length(x))=0;
else x(1,length(y))=0;
end
nx=x-mean(x);
ny=y-mean(y);
%Doplnenie vektorov o potrebny pocet 0 a vytvorenie matice
a=rem(length(nx),l);
a=(1-(a/l))*l;
a=length(x)+a;
nx(1,a)=0;
rx=reshape(nx,l,[]);rx=rx';
ny(1,a)=0;
ry=reshape(ny,l,[]);ry=ry';
% Overlap pre X,Y
b=round(l*(overlap/100));
nx=circshift(nx,[0 -b]);
ox=reshape(nx,l,[]);ox=ox';
ny=circshift(ny,[0 -b]);
oy=reshape(ny,l,[]);oy=oy';
x=horzcat(rx,ox);
y=horzcat(ry,oy);
w=hann(l*2);w=w';
xwt(length(x),l*2)=0;ywt=xwt;
for i=1:length(x)
    xw=w.*x(i,:); yw=w.*y(i,:);
    xwt(i,:)=abs(fft(xw));ywt(i,:)=abs(fft(yw));
end
dim=size(xwt);
for i=1:dim(1);
    for j=1:(l*2);
        d=10*log10(xwt(i,j));
        if d<-40;
```

```

        xwt(i,j)=-40;
    else xwt(i,j)=d;
    end
end
end
for i=1:dim(1);
    for j=1:(l*2);
        d=10*log10(ywt(i,j));
        if d<-40;
            ywt(i,j)=-40;
        else ywt(i,j)=d;
        end
    end
end
end

ul=64;
ll=2;
uh=128;
lh=65;
xlow=xwt(:,ll:ul);xlow=xlow';
ylow=ywt(:,ll:ul);ylow=ylow';
xhigh=xwt(:,lh:uh);xhigh=xhigh';
yhigh=ywt(:,lh:uh);yhigh=yhigh';
etl=1/(ul-ll+1)*sum(abs(ylow-xlow));
eth=1/(uh-lh+1)*sum(abs(yhigh-xhigh));
ecl=1/(ul-ll+1)*sum(ylow)-1/(ul-ll+1)*sum(xlow);
ech=1/(uh-lh+1)*sum(yhigh)-1/(uh-lh+1)*sum(xhigh);
for i=1:dim(1);
    for j=1:uh-lh;
        yl(j,i)=ylow(j,i)-ecl(i);
    end
    for j=1:uh-ul;
        yh(j,i)=yhigh(j,i)-ech(i);
    end
end
end
efl=1/(ul-ll+1)*sum(abs(yl-xlow));
efh=1/(uh-lh+1)*sum(abs(yh-xhigh));
p1=mean(ecl);
p2=mean(ech);
p3=mean(efl);
q=(1+exp(-4.29+(p1*0.15)+(p2*0.0243)+(p3*0.449)))^-1;
fid = fopen('q.txt','a+');
fseek(fid,0,'eof');
fprintf(fid,'%f',q);
fclose(fid);
end

```